scl

Danny Calegari

1991 *Mathematics Subject Classification.* Primary 20J05, 57M07;
Secondary 20F12, 20F65, 20F67, 37E45, 37J05, 90C05

Key words and phrases. stable commutator length, bounded cohomology, rationality, Bavard's Duality Theorem, hyperbolic groups, free groups, Thurston norm, Bavard's Conjecture, rigidity, immersions, causality, group dynamics, Markov chains, central limit theorem, combable groups, finite state automata

Supported in part by NSF Grants DMS-0405491 and DMS-0707130.

Abstract. This book is a comprehensive introduction to the theory of stable commutator length, an important subfield of quantitative topology, with substantial connections to 2-manifolds, dynamics, geometric group theory, bounded cohomology, symplectic topology, and many other subjects. We use constructive methods whenever possible, and focus on fundamental and explicit examples. We give a self-contained presentation of several foundational results in the theory, including Bavard's Duality Theorem, the Spectral Gap Theorem, the Rationality Theorem, and the Central Limit Theorem. The contents should be accessible to any mathematician interested in these subjects, and are presented with a minimal number of prerequisites, but with a view to applications in many areas of mathematics.

Preface

The historical roots of the theory of bounded cohomology stretch back at least as far as Poincaré [**167**] who introduced rotation numbers in his study of circle diffeomorphisms. The Milnor–Wood inequality [**154, 204**] as generalized by Sullivan [**193**], and the theorem of Hirsch–Thurston [**109**] on foliated bundles with amenable holonomy groups were also landmark developments.

But it was not until the appearance of Gromov's seminal paper [**97**] that a number of previously distinct and isolated phenomena crystallized into a coherent subject. In [**97**] and in [**98**] Gromov indicated how many important or delicate geometric and algebraic properties of groups could be encoded and (in principle) recovered from their bounded cohomology. The essence of bounded cohomology is that it is a functor from the category of groups and homomorphisms to the category of normed vector spaces and norm-decreasing linear maps. Theorems in bounded cohomology can be restated as algebraic or topological inequalities; rigidity phenomena arise when equality is achieved (see e.g. [**31, 93, 149, 45**]).

A certain amount of activity followed; for example, the papers [**6, 27, 115, 150**] contain significant new ideas and advanced the subject. But there is a sense in which the promise of the field as suggested by Gromov has not been realized. One major shortcoming is the lack of adequate tools for computing or extracting meaningful information. There are at least two serious technical problems:

(1) the failure of the standard machinery of homological algebra (e.g. spectral sequences) to carry over to the bounded cohomology context in a straightforward way
(2) the fact that in the cases of most interest (e.g. hyperbolic groups) bounded cohomology is usually so big as to be unmanageable

Monod's monograph [**157**] addresses in a very useful way some of the most serious shortcomings of the subject by largely restricting attention to *continuous* bounded cohomology in contexts where this restriction is most informative. Burger and Monod (see especially [**33**] and [**34**]) developed the theory of continuous bounded cohomology into a powerful tool, which is of most value to people working in ergodic theory or the theory of lattices (especially in higher-rank) but is less useful for people whose main concern is the bounded cohomology of discrete groups (although Theorem 2 from [**34**] is an exception).

To get an idea of the state of the subject *ca.* 2000, we quote an excerpt from Burger–Monod [**35**], p. 19:

> Although the theory of bounded cohomology has recently found many applications in various fields ... for discrete groups it remains scarcely accessible to computation. As a matter of fact,

> almost all known results assert either a complete vanishing or
> yield intractable infinite dimensional spaces.

It is therefore a firm goal of this monograph to try to present results in terms which are concrete and elementary. We pay a great deal of attention to the case of free and surface groups, and present efficient algorithms to compute numerical invariants, whenever possible.

It is always hard for an outsider (or even an insider) to get an accurate idea of the critical (internal) questions or conjectures in a given field, whose resolution would facilitate significant progress, and of how the field does or might connect to other threads in mathematics. This monograph has a number of modest aims:

(1) to restrict attention and focus to a subfield (namely stable commutator length) which already has a number of useful and well-known applications to a wide range of geometrical contexts
(2) to carefully expose a number of foundational results in a way which should be accessible to any mathematician interested in the subject, and with a minimal number of prerequisites
(3) to develop a number of "hooks" into the subject which invite contributions from mathematicians and mathematics in what might at first glance appear to be unrelated fields (representation theory, computer science, combinatorics, etc.)
(4) to highlight the importance of hyperbolic groups in general, and free groups in particular as a critical case for understanding certain basic phenomena
(5) to give an exposition of some of my own work, and that of my collaborators, especially that part devoted to the "foundations" of the subject

Recently, there has been an outburst of activity at the intersection of low-dimensional bounded cohomology, low-dimensional dynamics, and symplectic topology (e.g. [**71, 73, 86, 169, 170, 174**], and so forth). I have done my best to discuss some of the highlights of this interaction, but I am not competent to delve into it too deeply.

Danny Calegari

Acknowledgments

Thanks to Marc Burger, Jean-Louis Clerc, Matthew Day, Benson Farb, David Fisher, Koji Fujiwara, Ilya Kapovich, Dieter Kotschick, Justin Malestein, Fedor Manin, Curt McMullen, Geoff Mess, Assaf Naor, Andy Putman, Pierre Py, Peter Sarnak, Alden Walker, Anna Wienhard, Dave Witte-Morris and Dongping Zhuang for additions and corrections. Special thanks to Jason Manning for extensive comments (especially on Chapters 1 and 2), and for many useful conversations about numerous technical points. Extra-special thanks to Shigenori Matsumoto for his meticulous refereeing, which led to vast improvements throughout the book. Finally, thanks to Tereez for her patience.

Contents

CHAPTER 1

Surfaces

In this chapter we present some of the elements of the geometric theory of 2-dimensional (bounded) homology in an informal way. The main purpose of this chapter is to standardize definitions, to refresh the reader's mind about the relationship between 2-dimensional homology classes and maps of surfaces, and to compute the Gromov norm of a hyperbolic surface with boundary. All of this material is essentially elementary and many expositions are available; for example, [**10**] covers this material well.

We start off by discussing maps of surfaces into topological spaces. One way to study such maps is with linear algebra; this way leads to homology. The other way to study such maps is with group theory; this way leads to the fundamental group and the commutator calculus. These points of view are reconciled by Hopf's formula; a more systematic pursuit leads to rational homotopy theory.

1.1. Triangulating surfaces

A *surface* is a topological space (usually Hausdorff and paracompact) which is locally two dimensional. That is, every point has a neighborhood which is homeomorphic to the plane, usually denoted by $\mathbb{R}^2$.

1.1.1. The plane. It is unfortunate in some ways that the standard way to refer to the plane emphasizes its product structure. This product structure is topologically unnatural, since it is defined in a way which breaks the natural topological symmetries of the object in question. This fact is thrown more sharply into focus when one discusses more rigid topologies.

EXAMPLE 1.1 (Zariski topology). The product topology on two copies of the affine line with its Zariski topology is *not* typically the same as the Zariski topology on the affine plane. A closed set in $\mathbb{R}^1$ with the Zariski topology is either all of $\mathbb{R}$, or a finite collection of points. A closed set in $\mathbb{R}^2$ with the product topology is therefore either all of $\mathbb{R}^2$, or a finite union of horizontal and vertical lines and isolated points. By contrast, closed sets in the Zariski topology in $\mathbb{R}^2$ include circles, ovals, and algebraic curves of every degree.

Part of the bias is biological in origin:

EXAMPLE 1.2 (Primary visual cortex). The primary visual cortex of mammals (including humans), located at the posterior pole of the occipital cortex, contains neurons hardwired to fire when exposed to certain spatial and temporal patterns. Certain specific neurons are sensitive to stimulus along specific orientations, but in primates, more cortical machinery is devoted to representing vertical and horizontal than oblique orientations (see for example [**58**] for a discussion of this effect).

The correct way to discuss the plane is in terms of the separation properties of its 1-dimensional subsets. The foundation of many such results is the Jordan curve theorem, which says that there is essentially only one way to embed a circle in the plane, up to reparameterization and ambient homeomorphism. Moore [**158**] gave the first "natural" topological definition of the plane, in terms of separation properties of continua. Once this is understood, one is led to study the plane and other surfaces by cutting them up into simple pieces along 1-dimensional continua. The typical way to perform this subdivision is combinatorially, giving rise to triangulations.

1.1.2. Triangulations and homology. Every topological surface can be triangulated in an essentially unique way, up to subdivision (Radó [**176**]). Here by a *triangulation*, we mean a description of the surface as a simplicial complex built from countably many 2-dimensional simplices by identifying edges in pairs (note that a simplicial complex is topologized with the weak topology, so that every compact subset of a surface S meets only finitely many triangles).

Conversely, if we let $\coprod_i \Delta_i$ be a countable disjoint union of triangles, and glue the edges of the Δ_i in pairs, the result is a simplicial complex K. Every point in the interior of a face or an edge has a neighborhood homeomorphic to $\mathbb{R}^2$, by the gluing condition. Every vertex has a neighborhood homeomorphic to the open cone on its link. Each such link is a 1-manifold, and is therefore either homeomorphic to S^1 or to $\mathbb{R}$. It follows that the complex K is a surface if and only if the link of every vertex is *compact*.

If there are only finitely many triangles, every such identification gives rise to a surface. Otherwise, we need to impose the condition that each vertex in the quotient space is in the image of only finitely many triangles, so that the link of this vertex is compact.

REMARK 1.3. It is worth looking more closely at the set of all possible ways in which a given surface can be triangulated. Any two triangulations τ, τ' (given up to isotopy) of a fixed surface S are related by a finite sequence of local moves and their inverses. These moves are of two kinds: the *1–3 move*, and the *2–2 move*, illustrated in Figure 1.1. Only the 1–3 move and its inverse change the number of vertices in a triangulation, and

FIGURE 1.1. The 1–3 and the 2–2 moves

therefore these moves cannot be dispensed with entirely. However, it is an important fact that any two triangulations τ, τ' of the same surface S with the *same* number of vertices are related by 2–2 moves *alone*.

In fact, somewhat more than this is true. Define a *cellulation* of a surface to be a decomposition of the surface into polygonal disks (each with at least 3 sides). Associated to a surface S and a discrete collection P of points in S there is a natural cell complex $A(S,P)$ with one cell for each cellulation of S whose vertex set is exactly P, and with the property that one cell is in the boundary of another if one cellulation is obtained from the other by adding extra edges as diagonals in some of the polygons. In $A(S,P)$, the vertices correspond to the triangulations of S with vertex set exactly P, and the edges correspond to pairs of triangulations related by 2–2 moves. Hatcher [**105**] proves not only that $A(S,P)$ is connected, but that it is *contractible*.

The combinatorial view of a surface as a union of triangles gives rise to a fundamental relationship between surfaces and 2-dimensional homology.

EXAMPLE 1.4 (integral cycles). Let X be a topological space, and let $\alpha \in H_2(X)$ be an integral homology class. The class α is represented (possibly in many different ways) by an integral 2-cycle A. By the definition of a 2-cycle, there is an expression

$$A = \sum_i n_i \sigma_i$$

where each $n_i \in \mathbb{Z}$, and each σ_i is a singular 2-simplex; i.e. a continuous map $\sigma_i : \Delta^2 \to X$ where Δ^2 is the standard 2-simplex. By allowing repetitions of the σ_i, we can assume that each n_i is ± 1.

Since A is a cycle, $\partial A = 0$. That is, for each σ_i and for each singular 1-simplex e which is a face of some σ_i, the signed sum of copies of e appearing in the expression $\sum_i n_i \partial \sigma_i$ is 0. It follows that each such e appears an even number of times with opposite signs. This lets us choose a pairing of the faces of the σ_i so that each pair of faces contributes 0 in the expression for ∂A.

Build a simplicial 2-complex K by taking one 2-simplex for each σ_i, and gluing the edges according to this pairing. Since the number of simplices is finite, and edges are glued in pairs, the result is a topological surface S (note that S need not be connected). Each simplex of K can be oriented compatibly with the sign of the coefficient of the corresponding singular simplex σ_i, so the result is an *oriented* surface. The maps σ_i induce a map from the simplices of K into X, and the definition of the gluing implies that these maps are compatible on the edges of the simplices. We obtain therefore an induced continuous map $f_A : S \to X$. Since S is closed and oriented, there is a fundamental class $[S] \in H_2(S)$, and by construction we have

$$(f_A)_*([S]) = [A] = \alpha$$

In words, *elements of $H_2(X)$ are represented by maps of closed oriented surfaces into X.*

REMARK 1.5. One can also consider homology with rational or real coefficients. Every rational chain has a finite multiple which is an integral chain, so if one is prepared to consider "weighted" surfaces mapping to X, the discussion above suffices. We think of $H_2(X;\mathbb{Q})$ as a subset of $H_2(X;\mathbb{R})$ by using the natural isomorphism $H_2(X;\mathbb{Q}) \otimes \mathbb{R} = H_2(X;\mathbb{R})$. Suppose $\alpha \in H_2(X;\mathbb{Q})$ is represented by a real 2-cycle $A = \sum r_i \sigma_i$. Then for any $\epsilon > 0$ there exists a *rational* 2-cycle $A' = \sum r'_i \sigma_i$ (i.e. with the same support as A) such that the following are true:

(1) The cycles A and A' are homologous (hence $[A'] = \alpha$)
(2) There is an inequality $\sum_i |r_i - r'_i| < \epsilon$

To see this, let V denote the abstract vector space with basis the σ_i. There is a natural map $\partial : V \to C_1(X) \otimes \mathbb{R}$. Since ∂ is defined over $\mathbb{Q}$, the kernel $\ker(\partial)$ is a *rational* subspace of V. There is a further map $h : \ker(\partial) \to H_2(X;\mathbb{R}) = H_2(X;\mathbb{Q}) \otimes \mathbb{R}$. This map is also defined over $\mathbb{Q}$, and therefore $h^{-1}(\alpha)$ is a rational subspace of V (and therefore rational points are dense in it). Since A is in $h^{-1}(\alpha)$, it can be approximated arbitrarily closely by a rational cycle A' also in $h^{-1}(\alpha)$.

1.1.3. Topological classification of surfaces. For simplicity, in this section we consider only connected surfaces.

Closed surfaces are classified by Euler characteristic and whether or not they are orientable. For each non-negative integer g, there is a unique (up to homeomorphism) closed orientable surface with Euler characteristic $2-2g$. The number g is called the *genus* of S, denoted genus(S).

For each positive integer n, there is a unique (up to homeomorphism) closed non-orientable surface with Euler characteristic $2-n$.

Example 1.6 (closed surfaces). The sphere is the unique closed surface with $\chi(\text{sphere}) = 2$ and the torus is the unique closed orientable surface with $\chi(\text{torus}) = 0$. The projective plane is the unique closed surface with $\chi(\text{projective plane}) = 1$. For the sake of notation, we abbreviate these surfaces by S^2, T, P. Every closed surface may be obtained from these by the connect sum operation, denoted #. This operation is commutative and associative, with unit S^2, and satisfies

$$T\#P = P\#P\#P$$

Moreover, every other relation for # is a consequence of this one.

Euler characteristic is subadditive under connect sum, and satisfies

$$\chi(S_1\#S_2) = \chi(S_1) + \chi(S_2) - 2$$

A closed surface S is non-orientable if and only if P appears as a summand in some (and therefore any) expression of S as a sum of T and P terms.

If S is an oriented surface, we denote the same surface with opposite orientation by $\overline{S}$. We say that a topological surface is of *finite type* if it is homeomorphic to a closed surface minus finitely many points. If T is closed, and there is an inclusion $i : S \to T$ so that $T - i(S)$ is finite, then

$$\chi(S) = \chi(T) - \text{card}(T - i(S))$$

(here card denotes cardinality). Moreover, S is orientable if and only if T is.

1.1.4. Surfaces with boundary. A *surface with boundary* is a (Hausdorff, paracompact) topological space for which every point has a neighborhood which is either homeomorphic to $\mathbb{R}^2$ or to the closed half-space $\{(x, y) \in \mathbb{R}^2 \mid y \geq 0\}$. Points with neighborhoods homeomorphic to $\mathbb{R}^2$ are *interior points*, and the others are *boundary points*. Surfaces with boundary can be triangulated in such a way that the triangulation induces a triangulation (by 1-dimensional simplices) of the boundary. We denote the set of interior points of S by int(S), and the set of boundary points by ∂S.

If S is a surface with boundary, the *double* of S, denoted DS, is the surface obtained from $S \coprod \overline{S}$ by identifying ∂S with $\partial\overline{S}$. Note that $\overline{S}$ is only distinguished from S if S is oriented, in which case the double is also oriented. We say S is of finite type if DS is. Note that in this case, DS may be obtained from a closed surface DT which is the double of a compact surface with boundary T by removing finitely many points. If this happens, we can always assume S is obtained from T by removing finitely many points. Note that some of these points may be contained in ∂T.

Genus is not a good measure of complexity for surfaces with boundary: $-\chi$ is better, in the sense that there are only finitely many homeomorphism types of connected compact surface for which $-\chi$ is less than or equal to any given value.

1.1.5. Fundamental group and commutators. Let S be an oriented surface of finite type. If S has genus g and $p > 0$ punctures, $\pi_1(S)$ is free of rank $2g + p - 1$, and similarly if S is compact with p boundary components.

If S is closed of genus g, then S can be obtained by gluing the edges of a $4g$-gon in pairs, and one obtains the "standard" presentation of π_1:

$$\pi_1(S) = \langle a_1, b_1, \ldots, a_g, b_g \mid [a_1, b_1] \cdots [a_g, b_g] \rangle$$

A closed surface is obtained from a surface with one boundary component by gluing on a disk. If S has genus g with one boundary component, $\pi_1(S)$ is free with generators $a_1, b_1, \ldots, a_g, b_g$ and ∂S represents the conjugacy class of the element $[a_1, b_1] \cdots [a_g, b_g]$.

Let X be a topological space, and let α_i, β_i be elements in $\pi_1(X)$ such that there is an identity

$$[\alpha_1, \beta_1] \cdots [\alpha_g, \beta_g] = \mathrm{id} \in \pi_1(X)$$

There is an induced map $\pi_1(S) \to \pi_1(X)$ sending each $a_i \to \alpha_i$ and $b_i \to \beta_i$. Thinking of S as the quotient of a polygon P with $4g$ sides glued together in pairs, this defines a map $\partial P \to X$ whose image is null homotopic in X, and therefore this map extends to a map $S \to X$. The homology class of the image of the fundamental class $[S]$ depends on the particular expression involving the α_i, β_i. Moreover, two different choices of the extension $\partial P \to X$ to P differ by a pair of maps of P which agree on the boundary; these maps sew together to define a map $S^2 \to X$ defining an element of $\pi_2(X)$. In words, *identities in the commutator subgroup of* $\pi_1(X)$ *correspond to homotopy classes of maps of closed orientable surfaces into* X*, up to elements of* $\pi_2(X)$.

In the relative case, let $\gamma \in \pi_1(X)$ be a conjugacy class represented by a loop $l_\gamma \subset X$. If γ has a representative in the commutator subgroup $[\pi_1(X), \pi_1(X)]$ then we can write

$$[\alpha_1, \beta_1] \cdots [\alpha_g, \beta_g] = \gamma \in \pi_1(X)$$

Let S be a genus g surface with one boundary component. S is obtained from a $(4g + 1)$-gon P by identifying sides in pairs. Choose loops in X representing the elements $\gamma, \alpha_i, \beta_i$ and let $f : \partial P \to X$ be defined by sending the edges of P to loops in X by $a_i \to \alpha_i, b_i \to \beta_i$, and the free edge to γ. By construction, f factors through the quotient map $\partial P \to S$ induced by gluing up all but one of the edges. Moreover, by hypothesis, $f(\partial P)$ is null-homotopic in X. Hence f can be extended to a map $f : S \to X$ sending ∂S to γ.

In words, *loops corresponding to elements of* $[\pi_1(X), \pi_1(X)]$ *bound maps of oriented surfaces into* X.

1.1.6. Hopf's formula. The two descriptions above of (relative) maps of surfaces, in terms of homology and in terms of fundamental group, are related by Hopf's formula.

Let X be a topological space. If $\pi_2(X)$ is nontrivial, we can attach 3-cells to X to kill π_2 while keeping π_1 fixed. If X' is the result, then $H_2(X'; \mathbb{Z})$ can be identified with the group homology $H_2(\pi_1(X); \mathbb{Z})$, by the relative Hurewicz theorem.

We let $G = \pi_1(X)$. Suppose we have a description of G as a quotient of a free group:

$$0 \to R \to F \to G \to 0$$

where F is free. Every map from a closed oriented surface S into X' is associated to a product of commutators $[a_1, b_1] \cdots [a_g, b_g]$ which is equal to 0 in G. A choice of word in F for each element a_i, b_i in such an expression determines an element of $R \cap [F, F]$. A substitution $a'_i = a_i r$ where $r \in R$ changes the result by an element of $[F, R]$, since

$$[ar, b] = [ara^{-1}, aba^{-1}][a, b]$$

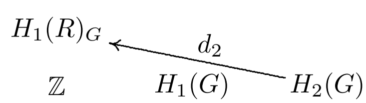
By the discussion above, there is a surjective homomorphism from the Abelian group $(R \cap [F, F])/[F, R]$ to $H_2(G)$.

Hopf's formula says this map is an isomorphism:

THEOREM 1.7 (Hopf's formula [**155**]). *Let G be a group written as a quotient $G = F/R$ where F is free. Then*

$$H_2(G) = (R \cap [F, F])/[F, R]$$

One quick way to see this is to use spectral sequences. This argument is short but a bit technical, and can be skipped by the novice, since the result will not be used elsewhere in this book. The extension $R \to F \to G$ defines a spectral sequence (the Hochschild–Serre spectral sequence [**110**]) whose $E^2_{n,0}$ term is $H_n(G)$ and whose $E^2_{0,1}$ term is $H_1(R)_G$, the quotient of $H_1(R)$ by the conjugation action of G. Since $H_1(R) = R/[R, R]$, we conclude that $H_1(R)_G$ is equal to $R/[F, R]$. Let $d_2 : E^2_{2,0} \to E^2_{0,1}$ be the differential connecting $H_2(G)$ to $R/[F, R]$:



Then there is an exact sequence

$$H_2(F) \to H_2(G) \to R/[F, R] \to H_1(F) \to H_1(G) \to 0$$

Since F is free, $H_2(F) = 0$ and therefore $H_2(G)$ is identified with the kernel of the map $R/[F, R] \to H_1(F)$. But the kernel of $F \to H_1(F)$ is exactly $[F, F]$, so we obtain Hopf's formula.

1.2. Hyperbolic surfaces

1.2.1. Conformal structures. A conformal structure on a surface is an atlas of charts for which the induced transition maps are angle-preserving. We do not require these maps to preserve the sense of the angles, so that non-orientable surfaces may still admit conformal structures. Orientable surfaces with conformal structures on them are synonymous with Riemann surfaces.

EXAMPLE 1.8 (conformal surfaces by cut-and-paste). A Euclidean polygon P inherits a natural conformal structure from the Euclidean plane, which we denote by $\mathbb{E}^2$. Isometries of $\mathbb{E}^2$ preserve the conformal structure, and therefore induce a natural conformal structure on any Euclidean surface. If S is obtained by gluing a locally finite collection of Euclidean polygons by isometries of the edges, the resulting surface is Euclidean away from the vertices, where there might be an angle deficit or surplus. If v is a vertex which has a cone angle of $r\pi$, we can develop the complement of v locally to the complement of the origin in $\mathbb{E}^2$. If we think of $\mathbb{E}^2$ as $\mathbb{C}$, and compose this developing map with the map $z \to z^{2/r}$ the result extends over v and defines a conformal chart near v which is compatible with the conformal charts on nearby points.

Let S be an arbitrary triangulated surface. By taking each triangle to be an equilateral Euclidean triangle with side length 1, and all gluing maps between edges to be isometries, we see that every surface can be given a conformal structure.

EXAMPLE 1.9 (Belyi's Theorem [**9**]). Belyi proved that a non-singular algebraic curve X is conformally equivalent to a surface obtained by gluing black and white equilateral triangles as above in a checkerboard pattern (i.e. so that no two triangles of the same color share an edge) if and only if X can be defined over an algebraic number field (i.e. a finite algebraic extension of $\mathbb{Q}$).

Such a description defines a map $X \to \mathbb{P}^1$, by taking the black triangles to the upper half space and the white triangles to the lower half space; this map is algebraic, and unramified except at $0, 1, \infty$. The preimage of the interval $[0, 1]$ is a bipartite graph on X, which Grothendieck called a "dessin d'enfant" (child's drawing; see [**100**]). The point of this construction is that the algebraic curve X can be recovered from the combinatorics and topology of the diagram. The Galois group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on the set of all dessins, and gives unexpected topological insight into this fundamental algebraic object.

A conformal structure on S induces a tautological conformal structure on $\overline{S}$. We say that a conformal structure on S is *conformally finite* if it is conformally equivalent to a closed surface minus finitely many points. Every surface of finite type admits a conformal structure which is conformally finite.

The classical *Uniformization Theorem* for Riemann surfaces says that any surface S with a conformal structure admits a complete Riemannian metric of constant curvature in its conformal class, which is unique up to similarity (note that this theorem is also valid for conformal surfaces of *infinite* type).

1.2.2. Conformal structures on surfaces with boundary. Let S be a surface with boundary. We say that a conformal structure on S is given by a conformal structure on DS which induces the same conformal structures on the interiors of S and $\overline{S}$ by inclusion, after composing with the tautological identification of S and $\overline{S}$. A surface with boundary S is said to be of finite type if DS is of finite type, and a conformal structure on S is *conformally finite* if DS is conformally finite.

If S admits a conformally finite conformal structure, we *define*

$$\chi(S) = \frac{1}{2}\chi(DS)$$

Note that this may not be an integer, but always takes values in $\frac{1}{2}\mathbb{Z}$.

If T is compact with boundary, and there is an inclusion $i : S \to T$ so that $T - i(S)$ is finite, then

$$\chi(S) = \chi(T) - \text{card}(\text{int}(T) - i(\text{int}(S))) - \frac{1}{2}\text{card}(\partial T - i(\partial S))$$

1.2.3. Hyperbolic surfaces. A Riemannian metric on a surface S is said to be *hyperbolic* if it has curvature -1 everywhere. A conformally finite surface admits a unique compatible hyperbolic metric which is complete of finite area if and only if $\chi(S) < 0$. The *Gauss–Bonnet Theorem* says that for any closed Riemannian surface S there is an equality

$$\int_S K = 2\pi\chi(S)$$

where K is the sectional curvature on S.

If S is hyperbolic, we obtain an equality

$$\text{area}(S) = -2\pi\chi(S)$$

A conformally finite surface S with boundary admits a unique hyperbolic structure for which ∂S is totally geodesic if and only if $\chi(S) < 0$. For, by definition, $\chi(DS) < 0$ and therefore DS admits a unique complete finite area hyperbolic structure in its conformal class. If $i : DS \to DS$ is the involution which interchanges S and $\overline{S}$, then i preserves the conformal structure, and therefore it acts on DS as an isometry. It follows that the fixed point set, which can be identified with ∂S, is totally geodesic. Notice that with our definition of $\chi(S)$ the relation

$$\text{area}(S) = -2\pi\chi(S)$$

holds also for surfaces with boundary.

1.2.4. Straightening chains. Let Δ be a geodesic triangle in $\mathbb{H}^2$. The Gauss–Bonnet Theorem gives a straightforward relationship between the area of Δ and the sum of the interior angles:

$$\text{area}(\Delta) = \pi - \text{sum of interior angles of } \Delta$$

It follows that there is a fundamental inequality

$$\text{area}(\Delta) < \pi$$

A geodesic triangle in $\mathbb{H}^2$ is *semi-ideal* if some of its vertices lie at infinity, and *ideal* if all three vertices are at infinity. If we allow Δ to be semi-ideal above, the inequality becomes

$$\text{area}(\Delta) \leq \pi$$

with equality if and only if Δ is ideal.

Similar inequalities hold in every dimension; that is, for every dimension m there is a constant $c_m > 0$ such that every geodesic hyperbolic m-simplex has volume $\leq c_m$, with equality if and only if the simplex is ideal and regular (Haagerup and Munkholm [**101**]). Note that every ideal 2-simplex is regular.

A fundamental insight, due originally to Thurston, is that in a hyperbolic manifold M^m, a singular chain can be replaced by a (homotopic) chain whose simplices are all geodesic. Applying this observation to the fundamental class $[M]$ of M, and observing that there is an upper bound on the volume of a geodesic simplex in each dimension, we see that the complexity (in a suitable sense) of a chain representing $[M]$ can be *bounded from below* in terms of c_m and $\text{vol}(M)$. That is, one can use (hyperbolic) geometry to estimate the complexity of an *a priori* topological quantity. Technically, the right way to quantify the complexity of $[M]$ is to use bounded (co-)homology, which we will study in detail in Chapter 2.

Definition 1.10. Let M be a hyperbolic m-manifold, and let $\sigma : \Delta^n \to M$ be a singular n-simplex. Define the *straightening* σ_g of σ as follows. First, lift σ to a map from Δ^n to $\mathbb{H}^m$ which we denote by $\tilde{\sigma}$.

Let $v_0, \cdots, v_n$ denote the vertices of Δ^n. In the hyperboloid model of hyperbolic geometry, $\mathbb{H}^m$ is the positive sheet (i.e. the points where $x_{m+1} > 0$) of the hyperboloid $\|x\| = -1$ in $\mathbb{R}^{m+1}$ with the inner product

$$\|x\| = x_1^2 + x_2^2 + \cdots + x_m^2 - x_{m+1}^2$$

If $t_0, \cdots, t_n$ are barycentric co-ordinates on Δ^n, so that $v = \sum_i t_i v_i$ is a point in Δ^n, define

$$\tilde{\sigma}_g(v) = \frac{\sum_i t_i \tilde{\sigma}(v_i)}{-\| \sum_i t_i \tilde{\sigma}(v_i) \|}$$

and define σ_g to be the composition of $\tilde{\sigma}$ with projection $\mathbb{H}^m \to M$.

Since the isometry group of $\mathbb{H}^m$ acts on $\mathbb{R}^{m+1}$ linearly preserving the form $\| \cdot \|$, the straightening map $\sigma \to \sigma_g$ is well-defined, and independent of the choice of lift.

Let M be a hyperbolic manifold. Define

$$\mathrm{str} : C_*(M) \to C_*(M)$$

by setting $\mathrm{str}(\sigma) = \sigma_g$, and extending by linearity.

By composing a linear homotopy in $\mathbb{R}^{m+1}$ with radial projection to the hyperboloid, one sees that there is a chain homotopy between str and the identity map.

1.2.5. The Gromov norm. We now return to hyperbolic surfaces. Let S be conformally finite, possibly with boundary. If S is closed and oriented, the *fundamental class* of S, denoted $[S]$, is the generator of $H_2(S, \partial S)$ which induces the orientation on S.

Definition 1.11. Define the L^1 norm, also called the *Gromov norm* of S, as follows. Consider the homomorphism

$$i_* : H_2(S, \partial S; \mathbb{Z}) \to H_2(S, \partial S; \mathbb{R})$$

induced by inclusion $\mathbb{Z} \to \mathbb{R}$, and by abuse of notation, let $[S]$ denote the image of the fundamental class. Let $C = \sum_i r_i \sigma_i$ represent $[S]$, where the coefficients r_i are real, and denote

$$\|C\|_1 = \sum_i |r_i|$$

Then set

$$\|[S]\|_1 = \inf_C \|C\|_1$$

The following lemma, while elementary, is very useful in what follows.

Lemma 1.12. *Let S be an orientable surface with p boundary components. If $p > 1$ then for any integer $m > 1$ with m and $p-1$ coprime there is an m-fold cyclic cover S_m with p boundary components, each of which maps to the corresponding component of ∂S by an m-fold covering.*

Proof. The inclusion $\partial S \to S$ induces a homomorphism $H_1(\partial S) \to H_1(S)$ whose kernel is 1-dimensional, and generated by the homology class represented by the union ∂S. In particular, if $p > 1$, then we can take $p - 1$ boundary components to be part of a basis for $H_1(S)$. Denote the images of the boundary components in $H_1(S)$ by $e_1, \cdots, e_p$, and let $e_1, \cdots, e_{p-1}$ be part of a basis for $H_1(S)$. If m and $p - 1$ are coprime, let $\alpha \in H^1(S; \mathbb{Z}/m\mathbb{Z}) = \mathrm{Hom}(H_1(S); \mathbb{Z}/m\mathbb{Z})$ satisfy $\alpha(e_i) = 1$ for $1 \le i \le p-1$. Then $\alpha(e_j)$ is primitive for all $1 \le j \le p$. The kernel of α defines a regular m-fold cover S_m with the desired properties. $\square$

REMARK 1.13. A surface with exactly one boundary component has no regular (nontrivial) covers with exactly one boundary component. However *irregular* covers with this property do exist. For example, let S be a genus one surface with one boundary component, so $\pi_1(S)$ is free on two generators a, b. Let $\phi : \pi_1(S) \to S_3$ be the permutation representation defined by $\phi(a) = (12)$ and $\phi(b) = (23)$. Then $\phi([a,b]) = (312)$, which is a 3-cycle. This representation determines a 3-sheeted (irregular) cover of S with one boundary component.

It is straightforward to generalize this example to show that every connected oriented surface with $\chi \leq 0$ admits a connected cover of arbitrarily large degree with the same number of boundary components.

THEOREM 1.14 (Gromov norm of a hyperbolic surface). *Let S be a compact orientable surface with $\chi(S) < 0$, possibly with boundary. Then*

$$\|[S]\|_1 = -2\chi(S)$$

PROOF. Let S be a surface of genus g with p boundary components, so that

$$\chi(S) = 2 - 2g - p$$

The surface S admits a triangulation with one vertex on each boundary component, and no other vertices. Any such triangulation has $4g + 3p - 4$ triangles. Figure 1.2 exhibits the case $g = 1$, $p = 2$. By Lemma 1.12, there is an m-fold cover S_m of S with p boundary components.

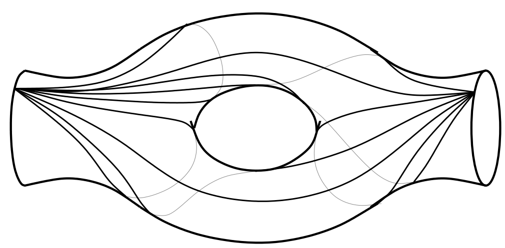


FIGURE 1.2. A triangulation of a surface with $g = 1, p = 2$ by 6 triangles

Since χ is multiplicative under covers, $\chi(S_m) = 2m - 2gm - mp$ and it can be triangulated with $p + m(4g + 2p - 4)$ triangles. Projecting this triangulation under the covering map $S_m \to S$ gives an integral chain representing $m[S]$ with L^1 norm equal to $p + m(4g + 2p - 4)$. Dividing coefficients by m and taking the limit as $m \to \infty$, we get

$$\|[S]\|_1 \leq -2\chi(S)$$

To obtain the other inequality, let C be any chain representing $[S]$. Then $\text{str}(C)$ has L^1 norm no greater than that of C, and also represents $[S]$. On the other hand, since every geodesic triangle has area $\leq \pi$, and $\text{area}(S) = -2\pi\chi(S)$, we obtain

$$\|[S]\|_1 \geq -2\chi(S)$$

□

REMARK 1.15. If $\chi(S) \geq 0$ then S admits a proper self map $f : S \to S$ of any degree. By pushing forward a chain under this map and dividing coefficients, one sees that $\|[S]\|_1 = 0$.

If X is any topological space and α is a class in $H_2(X;\mathbb{R})$ the Gromov norm of α, denoted $\|\alpha\|_1$, is the infimum of the L^1 norm over all (real valued) 2-cycles representing the homology class α. If α is rational, any real 2-cycle representing

α can be approximated in L^1 by a rational 2-cycle representing α. By multiplying through to clear denominators, some multiple $n\alpha$ can be represented by a map of a surface $S \to X$. For such a surface S, let $\chi^-(S)$ denote the Euler characteristic of the union of the non-spherical components of S. Then Theorem 1.14 implies that

$$\|\alpha\|_1 = \inf_S \frac{-2\chi^-(S)}{n(S)}$$

where the image of the fundamental class of S under the map $S \to X$ represents $n(S)\alpha$ in homology, and the infimum is taken over all maps of (possibly disconnected) closed oriented surfaces into X.

CHAPTER 2

Stable commutator length

Many natural problems in topology and geometric group theory can be formulated as a kind of *genus problem*. In the absolute version of this problem, one is given a space X and tries to find a surface in X with prescribed properties, of least genus. Examples of the kind of properties one wants for the surface are that it represent a given class in $H_2(X)$, that it is a Heegaard surface (in a 3-manifold), that $\pi_1(X)$ splits nontrivially over its image, that it is pseudoholomorphic, etc. In the relative version one is given X and a loop γ in X and tries to find a surface (again with prescribed properties) of least genus with boundary γ. In its purest form, the analogue of this second problem in group theory asks to determine the *commutator length* of an element in the commutator subgroup of a group, and it is this problem (or rather its stabilization) with which we are preoccupied in this chapter (we give precise definitions in § 2.1). We will use the algebraic and geometric language interchangeably in what follows; however our *methods* and *arguments* are mostly geometric.

There is a dual formulation of these problems, in terms of (bounded) cohomology and *quasimorphisms* — real-valued functions on a group which are additive, up to bounded error. This duality is expressed in the fundamental *Bavard Duality theorem* from [**8**], which gives a precise relationship between (stable) commutator length and bounded cohomology, and reconciles the homotopy theoretic and the (co)-homological points of view of surfaces and the genus problem. The main goal of this chapter is to give a self-contained exposition of this fundamental result and some generalizations, including all the necessary background and details. Our aim is to keep the presentation elementary wherever possible, although certain arguments are streamlined by using the language of abstract functional analysis.

In many places we follow Bavard's original paper [**8**], though occasionally our emphasis is different. We also enumerate and prove some useful properties of scl and bounded cohomology which are used in subsequent chapters.

2.1. Commutator length and stable commutator length

Definition 2.1. Let G be a group, and $a \in [G,G]$. The *commutator length* of a, denoted $\mathrm{cl}(a)$, is the least number of commutators in G whose product is equal to a.

By convention we define $\mathrm{cl}(a) = \infty$ for a not in $[G,G]$.

Definition 2.2. For $a \in [G,G]$, the *stable commutator length*, denoted $\mathrm{scl}(a)$, is the following limit:

$$\mathrm{scl}(a) = \lim_{n\to\infty} \frac{\mathrm{cl}(a^n)}{n}$$

For each fixed a, the function $n \to \mathrm{cl}(a^n)$ is non-negative and subadditive; hence this limit exists. If a is not in $[G,G]$ but has a power a^n which is, define $\mathrm{scl}(a) = \mathrm{scl}(a^n)/n$, and by convention define $\mathrm{scl}(a) = \infty$ if and only if a represents a nontrivial element in $H_1(G;\mathbb{Q})$.

REMARK 2.3. Computing commutator length is almost always difficult, even in finite groups. Ore [**164**] famously conjectured in 1951 that every element of a finite non-cyclic simple group is a commutator, and proved his conjecture for alternating groups A_n where $n \ge 5$. After receiving considerable attention (see e.g. [**72, 121**]), Ore's conjecture was finally proved in 2008 by Liebeck–O'Brien–Shalev–Tiep [**135**].

Commutator length in free groups has been studied by many people, with effective (though inefficient) procedures for calculating commutator length first obtained by Edmunds [**68, 69**]. The use of geometric methods to study genus was pioneered by Culler [**59**]. Several authors ([**98, 99, 178**]) used minimal surface techniques to obtain estimates of commutator length under geometric hypotheses.

Thurston [**196**], studied the absolute genus problem in the context of *embedded* surfaces in 3-manifolds, and showed how a stabilization of this problem gives rise to a *norm* on homology with several remarkable properties. Gromov [**99**] also emphasized the importance of stabilization, and posed a number of very general problems about genus and stable genus, especially their interaction with negative curvature. Gromov further stressed the relationship between the stable genus problem and bounded cohomology, which he systematically introduced and studied in [**97**]. This connection was also studied by Matsumoto and Morita; the paper [**150**] describes a fundamental relationship between homological "filling" norms and the kernel of the natural map from bounded to ordinary cohomology.

The most important property of cl and scl is their *monotonicity* under homomorphisms:

LEMMA 2.4 (monotonicity). *Let $\varphi : G \to H$ be a homomorphism of groups. Then $\mathrm{scl}_H(\varphi(a)) \le \mathrm{scl}_G(a)$ for all $a \in G$ and similarly for* cl.

PROOF. The image of a commutator under a homomorphism is a commutator. It follows that both cl and scl are monotone decreasing. □

The following corollaries are immediate:

COROLLARY 2.5 (retraction). *Let $\varphi : G \to H$ be a monomorphism with a left inverse; i.e. there is $\psi : H \to G$ with $\psi \circ \varphi : G \to G$ the identity. Then*

$$\mathrm{scl}(\varphi(a)) = \mathrm{scl}(a)$$

for all $a \in G$.

COROLLARY 2.6 (characteristic). *The functions* cl *and* scl *are constant on orbits of* $\mathrm{Aut}(G)$.

REMARK 2.7. Corollary 2.6 is especially interesting when $\mathrm{Out}(G)$ is large.

For most interesting phenomena concerning scl, it suffices to restrict attention to *countable* groups, as the following Lemma shows.

LEMMA 2.8 (countable). *Let G be a group, and $a \in G$ an element. Then there is a countable subgroup $H < G$ containing a, such that $\mathrm{scl}_H(a) = \mathrm{scl}_G(a)$.*

PROOF. For each n, exhibit a^n as a product of $\mathrm{cl}(a^n)$ commutators in G, and let H_n be the subgroup generated by the elements appearing in these commutators. Then let H be the subgroup generated by $\cup_n H_n$. □

The algebraic definitions of cl and scl are almost useless for the purposes of computation. Products and powers of commutators satisfy many identities which at first glance might appear quite mysterious.

EXAMPLE 2.9 (Culler [**59**]). For any elements a, b in any group, there is an identity

$$[a,b]^3 = [aba^{-1}, b^{-1}aba^{-2}][b^{-1}ab, b^2]$$

These properties are often more clear from a geometric perspective (for instance, Example 2.9 is really just Remark 1.13 in disguise). Given a group G, one can construct a space X (for example, a CW complex) with $\pi_1(X) = G$. A conjugacy class $a \in G$ corresponds to a free homotopy class of loop γ in X. From the definitions and the discussion in § 1.1.5 it follows that the commutator length of a is the least genus of a surface with one boundary component mapping to X in such a way that the boundary represents the free homotopy class of γ, and the stable commutator length of a may be obtained by estimating the genus of surfaces whose boundary wraps multiple times around γ.

Once we have recast this problem in geometric terms, a number of facts become immediately apparent:

(1) genus is not multiplicative under coverings whereas Euler characteristic is
(2) there is no good reason to restrict attention to surfaces with exactly one boundary component

As in § 1.2.5, given a (not necessarily connected) compact oriented surface S, let $-\chi^-(S)$ denote the sum of $\max(-\chi(\cdot), 0)$ over the components of S. Given a space X and a loop $\gamma : S^1 \to X$ we say that a map $f : S \to X$ is *admissible* if there is a commutative diagram:

$$\begin{array}{ccc} \partial S & \xrightarrow{\quad i \quad} & S \\ {\scriptstyle \partial f}\downarrow & & \downarrow{\scriptstyle f} \\ S^1 & \xrightarrow[\quad \gamma \quad]{} & X \end{array}$$

Since S is oriented, the boundary of S inherits an orientation, and it makes sense to define the fundamental class $[\partial S]$ in $H_1(\partial S)$. Similarly, one has a fundamental class $[S^1] \in H_1(S^1)$. Define $n(S)$ by the formula

$$\partial f_*[\partial S] = n(S)[S^1]$$

Note that by orienting S appropriately, we can ensure that $n(S) \geq 0$. The number $n(S)$ is just the (total algebraic) *degree* of the map $\partial S \to S^1$ between oriented closed manifolds.

With this notation, one can give an intrinsically geometric definition of scl, which is contained in the following proposition.

PROPOSITION 2.10. *Let* $\pi_1(X) = G$, *and let* $\gamma : S^1 \to X$ *be a loop representing the conjugacy class of* $a \in G$. *Then*

$$\mathrm{scl}(a) = \inf_S \frac{-\chi^-(S)}{2n(S)}$$

where the infimum is taken over all admissible maps as above.

PROOF. An inequality in one direction is obvious: $\mathrm{cl}(a^n) \le g$ if and only if there is an admissible map $f : S \to X$, where S has exactly one boundary component and satisfies $n(S) = n$ and $2g - 1 = -\chi^-(S)$. Hence $\lim_n \mathrm{cl}(a^n)/n \ge \inf_S -\chi^-(S)/2n(S)$.

Conversely, suppose $f : S \to X$ is admissible. If S has multiple components, at least one of them S_i satisfies $-\chi^-(S_i)/2n(S_i) \le -\chi^-(S)/2n(S)$, so without loss of generality we can assume S is connected. Since $-\chi^-(\cdot)$ and $2n(\cdot)$ are both multiplicative under covers, we can replace S with any finite cover without changing their ratio, so we may additionally assume that S has $p \ge 2$ boundary components.

As in Lemma 1.12, we can find a finite cover $S' \to S$ of degree $N \gg 1$ such that S' also has p boundary components. Observe that $-\chi^-(S') = -N\chi^-(S)$ and $n(S') = Nn(S)$. We may modify S' by attaching 1-handles to connect up the different boundary components, and extend $\partial f'$ over these 1-handles by a trivial map to a basepoint of S^1. Adding a 1-handle increases genus by 1 and reduces the number of boundary components by 1, so it increases $-\chi^-$ by 1. The result of this is that we can find a new surface S'' with *exactly one* boundary component and a map f'' satisfying $-\chi^-(S'') = -\chi^-(S') + p - 1$ and $n(S'') = n(S')$. We estimate

$$\frac{-\chi^-(S'')}{2n(S'')} = \frac{p - 1 - N\chi^-(S)}{2Nn(S)}$$

Since S is arbitrary, and given S the number p is fixed but N may be taken to be as large as desired, the right hand side may be taken to be arbitrarily close to $\inf_S -\chi^-(S)/2n(S)$. On the other hand, since the genus of S'' may be chosen to be as large as desired, and since S'' has exactly one boundary component, we have $\mathrm{cl}(a^{n(S'')}) \le -\chi^-(S'')/2 + 1$. The proof follows. □

Notice that for any element a of infinite order, we have an inequality $\mathrm{scl}(a) \le \mathrm{cl}(a^n)/n - 1/2n$. It follows that *no* surface can realize the infimum of $\mathrm{cl}(a^n)/n$. On the other hand, it is entirely possible for a surface to realize the infimum of $-\chi^-(S)/2n(S)$. Such surfaces are sufficiently useful and important that they deserve to be given a name.

DEFINITION 2.11. A surface $f : S \to X$ realizing the infimum of $-\chi^-(S)/2n(S)$ is said to be *extremal*.

We will return to extremal surfaces in § 4.1.10.

At this point it is convenient to state and prove another proposition about the kinds of admissible surfaces we need to consider.

DEFINITION 2.12. An admissible map $f : S, \partial S \to X, \gamma$ is *monotone* if for every boundary component ∂_i of ∂S, the degree of $\partial f : \partial_i \to S^1$ has the same sign.

PROPOSITION 2.13. *Let S be connected with $\chi(S) < 0$, and let $f : S, \partial S \to X, \gamma$ be admissible. Then there is a* monotone *admissible map $f' : S', \partial S' \to X, \gamma$ with $-\chi^-(S')/2n(S') \le -\chi^-(S)/2n(S)$.*

PROOF. Each boundary component ∂_i of ∂S maps to S^1 with degree n_i (which may be positive, negative or zero), where $\sum_i n_i = n(S)$. If some n_i is zero, the image $f(\partial_i)$ is homotopically trivial in X, so we may reduce $-\chi^-$ by compressing ∂_i. Hence we may assume every n_i is nonzero.

If S is a planar surface, then since $\chi(S) < 0$, there is a finite cover of S with positive genus. If S is a surface with positive genus and negative Euler characteristic, there is a degree 2 cover $S' \to S$ such that each boundary component in S has exactly two preimages. Hence, after passing to a finite cover if necessary, we can assume that the boundary components ∂_i come in *pairs* with equal degrees n_i.

Now let N be the least common multiple of the $|n_i|$. Define ϕ as a function on the set of boundary components with values in $\mathbb{Z}/N\mathbb{Z}$ as follows. For each pair of boundary components ∂_i, ∂_j with $n_i = n_j$, define $\phi(\partial_i) = n_i$ and $\phi(\partial_j) = -n_i$. Then $\sum_i \phi(\partial_i) = 0$, so ϕ extends to a surjective homomorphism from $\pi_1(S)$ to $\mathbb{Z}/N\mathbb{Z}$. If S' is the cover associated to the kernel, then each component of $\partial S'$ has degree $\pm N$. Pairs of components for which the sign of the degree is opposite can be glued up (which does not affect χ or $n(\cdot)$) until all remaining components have degrees with the same signs. □

Consequently it suffices to take the infimum of $-\chi^-/2n$ over monotone surfaces to determine scl.

Remark 2.14. Note that the surface constructed in Proposition 2.13 is not merely monotone, but has the property that all boundary components map with the *same* degree.

2.2. Quasimorphisms

We now have two different definitions of stable commutator length: an algebraic definition and a (closely related) topological definition. It turns out that one can also give a *functional analysis* definition, couched not directly in terms of groups and elements, but dually in terms of certain kinds of functions on groups, namely *quasimorphisms*. This particular form of duality is known as *Bavard duality*; the precise statement of this duality is Theorem 2.70.

2.2.1. Definition.

Definition 2.15. Let G be a group. A *quasimorphism* is a function

$$\phi : G \to \mathbb{R}$$

for which there is a least constant $D(\phi) \geq 0$ such that

$$|\phi(ab) - \phi(a) - \phi(b)| \leq D(\phi)$$

for all $a, b \in G$. In words, *a quasimorphism is a real-valued function which is additive up to bounded error*. The constant $D(\phi)$ is called the *defect* of ϕ.

Example 2.16. Any bounded function is a quasimorphism. A quasimorphism has defect 0 if and only if it is a homomorphism.

Lemma 2.17. *Let S be a (possibly infinite) generating set for G. Let w be a word in the generators, representing an element of G. Let $|w|$ denote the length of w, and let w_i denote the ith letter. Then*

$$\Big|\phi(w) - \sum_{i=1}^{|w|} \phi(w_i)\Big| \leq (|w| - 1)D(\phi)$$

Proof. This follows from the defining property of a quasimorphism, the triangle inequality, and induction. □

The set of all quasimorphisms on a fixed group G is easily seen to be a (real) vector space; we denote this vector space by $\widehat{Q}(G)$. In anticipation of what is to come, we denote the space of (real-valued) bounded functions on G by $C_b^1(G)$, and observe that C_b^1 is a vector subspace of $\widehat{Q}$.

2.2.2. Antisymmetric and homogeneous quasimorphisms. Some quasimorphisms are better behaved than others.

DEFINITION 2.18. A quasimorphism ϕ is *antisymmetric* if

$$\phi(a^{-1}) = -\phi(a)$$

for all a. Any quasimorphism ϕ can be *antisymmetrized* $\phi \to \phi'$ by the formula

$$\phi'(a) = \frac{1}{2}(\phi(a) - \phi(a^{-1}))$$

LEMMA 2.19. *For any quasimorphism ϕ, the antisymmetrization ϕ' satisfies*

$$D(\phi') \le D(\phi)$$

PROOF. We calculate

$$D(\phi') = \sup_{a,b} |\phi'(ab) - \phi'(a) - \phi'(b)|$$

$$= \sup_{a,b} \frac{1}{2}|\phi(ab) - \phi(a) - \phi(b) - \phi(b^{-1}a^{-1}) + \phi(a^{-1}) + \phi(b^{-1})| \le D(\phi)$$

□

Observe that for any antisymmetric quasimorphism ϕ there is an inequality

$$|\phi([a,b])| = |\phi(aba^{-1}b^{-1}) - \phi(a) - \phi(b) - \phi(a^{-1}) - \phi(b^{-1})| \le 3D(\phi)$$

and in general (by Lemma 2.17), $|\phi(\prod_{i=1}^n [a_i, b_i])| \le (4n-1)D(\phi)$.

DEFINITION 2.20. A quasimorphism is *homogeneous* if it satisfies the additional property

$$\phi(a^n) = n\phi(a)$$

for all $a \in G$ and $n \in \mathbb{Z}$. Denote the vector space of homogeneous quasimorphisms on G by $Q(G)$.

LEMMA 2.21. *Let ϕ be a quasimorphism on G. For each $a \in G$, define*

$$\overline{\phi}(a) := \lim_{n\to\infty} \frac{\phi(a^n)}{n}$$

The limit exists, and defines a homogeneous quasimorphism. Moreover, for any $a \in G$ there is an estimate $|\overline{\phi}(a) - \phi(a)| \le D(\phi)$

PROOF. For each positive integer i, there is an inequality

$$|\phi(a^{2^i}) - 2\phi(a^{2^{i-1}})| \le D(\phi)$$

dividing by 2^i and applying the triangle inequality and induction, we see that for any $j < i$,

$$|\phi(a^{2^i})2^j/2^i - \phi(a^{2^j})| \le D(\phi)$$

so $\phi(a^{2^i})2^{-i}$ is a Cauchy sequence. Define $\overline{\phi}(a)$ to be the limit $\lim_{i\to\infty} \phi(a^{2^i})2^{-i}$ and observe that $|\overline{\phi}(a) - \phi(a)| \le D(\phi)$ for all a.

Since $\overline{\phi} - \phi$ is in C^1_b, we conclude that $\overline{\phi}$ is a quasimorphism. It remains to show that $\overline{\phi}$ is homogeneous. For any j, by the definition of $\overline{\phi}$ we have

$$|\overline{\phi}(a^j) - j\overline{\phi}(a)| = \lim_{i\to\infty} 2^{-i}|\phi(a^{j2^i}) - j\phi(a^{2^i})| \le \lim_{i\to\infty} (j-1)D(\phi) \cdot 2^{-i} = 0$$

where the last inequality follows from Lemma 2.17. □

REMARK 2.22. Since $|\overline{\phi}(a)-\phi(a)| \le D(\phi)$ for any element a, the triangle inequality implies that $D(\overline{\phi}) \le 4 \cdot D(\phi)$. In fact, a more involved argument (Lemma 2.58) will give a better estimate of the defect.

Homogeneous quasimorphisms are often easier to work with than ordinary quasimorphisms, but ordinary quasimorphisms are easier to construct. We use this averaging procedure to move back and forth between the two concepts. Note that a homogeneous quasimorphism is already antisymmetric, and that homogenization commutes with antisymmetrization.

REMARK 2.23. If ϕ takes values in some additive subgroup $R \subset \mathbb{R}$ then the antisymmetrization may take values in $\frac{1}{2}R$, and the homogenization may take arbitrary values in $\mathbb{R}$.

2.2.3. Commutator estimates. If ϕ is homogeneous, then

$$|\phi(aba^{-1}) - \phi(b)| = \frac{1}{n}|(\phi(ab^n a^{-1}) - \phi(b^n))| \le \frac{2D(\phi)}{n}$$

Hence ϕ is constant on conjugacy classes; i.e. *homogeneous quasimorphisms are class functions*. It follows that for any commutator $[a,b] \in G$ and any homogeneous quasimorphism ϕ we have an inequality

$$|\phi([a,b])| \le D(\phi)$$

In fact, this inequality is always *sharp*:

LEMMA 2.24 (Bavard, Lemma 3.6. [**8**]). *Let ϕ be a homogeneous quasimorphism on G. Then there is an equality*

$$\sup_{a,b} |\phi([a,b])| = D(\phi)$$

PROOF. First we show that we can write $a^{2n}b^{2n}(ab)^{-2n}$ as a product of n commutators. If $n=1$ this is just the identity

$$a^2ba^{-1}b^{-1}a^{-1} = a[a,b]a^{-1} = [a, aba^{-1}]$$

Also,

$$a^{2n}b^{2n}(ab)^{-2n} = a(a^{2n-1}b^{2n-1}(ba)^{-2n+1})a^{-1}$$

so it suffices to show that $a^{2n-1}b^{2n-1}(ba)^{-2n+1}$ can be written as a product of n commutators.

We proceed by induction, and assume we have proved this for $n \le m$. Then

$$\begin{aligned}[][a^{-2m+1}b^{-2m}a^{-2}, ab^{-1}a^{2m-1}] &= a^{-2m+1}b^{-2m}a^{-1}b^{-1}a^{2m+1}b^{2m+1}a^{-1} \\ &= a(a^{-2m}b^{-2m}a^{-1}b^{-1}a^{2m+1}b^{2m+1})a^{-1}\end{aligned}$$

By induction, and after interchanging a and b for a^{-1} and b^{-1}, the expression $a^{-2m}b^{-2m}$ can be written as a product of m commutators times $(a^{-1}b^{-1})^{2m}$. It follows that $(a^{-1}b^{-1})^{2m+1}a^{2m+1}b^{2m+1}$ can be written as a product of $m+1$ commutators, and the induction step is complete, proving the claim.

Now let a, b be chosen so that $|\phi(ab) - \phi(a) - \phi(b)| \ge D(\phi) - \epsilon$ for some small ϵ (to be chosen later). Since ϕ is homogeneous, for any n we have

$$|\phi((ab)^{2n}) - \phi(a^{2n}) - \phi(b^{2n})| \ge 2n(D(\phi) - \epsilon)$$

On the other hand, we have shown that $(ab)^{2n}$ can be expressed as a product of n commutators c_i (which depend on a and b) times $a^{2n}b^{2n}$. Hence by Lemma 2.17,

$$|\phi((ab)^{2n}) - \phi(a^{2n}) - \phi(b^{2n}) - \sum_{i=1}^{n} \phi(c_i)| \leq (n+1)D(\phi)$$

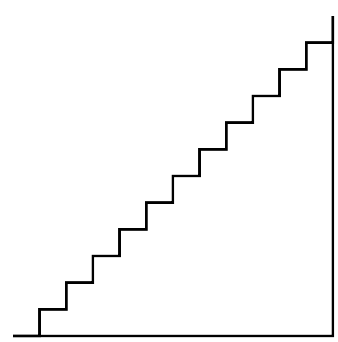
By the triangle inequality,

$$|\sum_{i=1}^{n} \phi(c_i)| \geq (n-1)D(\phi) - 2n\epsilon$$

Since $\phi(c_i) \leq D(\phi)$ for every commutator, taking n to be big, and then ϵ small compared to $1/n$, we see that some commutator c_i has $\phi(c_i)$ as close to $D(\phi)$ as we like. □

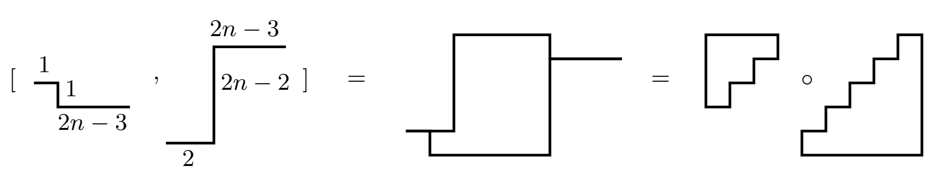
2.2.4. Graphical calculus. The argument that $a^{2n}b^{2n}(ab)^{-2n}$ can be written as a product of n commutators can be expressed more simply in the form of a graphical calculus.

A word w in F_2 determines a path in the square lattice $\mathbb{Z}^2$. Such a path corresponds to a reduced word if and only if it has no backtracking. It represents a commutator in F_2 if and only if it closes up to a loop. If one disregards basepoints, loops correspond to cyclic conjugacy classes of elements in $[F_2, F_2]$.



In this calculus, the word $a^{2n}b^{2n}(ab)^{-2n}$ is represented by the loop indicated in the figure. Note that this word is unreduced: there are two spurious backtracks, each of length 1. After removing these backtracks, one obtains a loop representing the word $a^{2n-1}b^{2n-1}(ba)^{-2n+1}$

Informally, the word $a^{2n-1}b^{2n-1}(ba)^{-2n+1}$ is a "staircase" of height $2n-1$. In this language, the induction step can be expressed as saying that a staircase of height $2n-1$ can be written as the product of a commutator with a staircase of height $2n-3$. Since a staircase of height 1 is just the commutator $[a, b]$, this completes the proof. This can be expressed graphically in the following way:



2.3. Examples

In this section we discuss some fundamental examples of quasimorphisms. These examples can all be generalized considerably, as we shall see in later Chapters.

2.3.1. de Rham quasimorphisms. The following construction is due to Barge–Ghys [**6**].

Let M be a closed hyperbolic manifold, and let α be a 1-form. Define a quasimorphism $q_\alpha : \pi_1(M) \to \mathbb{R}$ as follows. Choose a basepoint $p \in M$. For each $\gamma \in \pi_1(M)$, let L_γ be the unique oriented geodesic arc with both endpoints at p which as a based loop represents γ in $\pi_1(M)$. Then define

$$q_\alpha(\gamma) = \int_{L_\gamma} \alpha$$

If γ_1, γ_2 are two elements of $\pi_1(M)$, there is a geodesic triangle T whose oriented boundary is the union of $L_{\gamma_1}, L_{\gamma_2}, L_{\gamma_2^{-1}\gamma_1^{-1}}$. By Stokes' theorem we can calculate

$$q_\alpha(\gamma_1) + q_\alpha(\gamma_2) - q_\alpha(\gamma_1\gamma_2) = \int_T d\alpha$$

A geodesic triangle in a hyperbolic manifold has area at most π. It follows that the defect of q_α is at most $\pi \cdot \|d\alpha\|$.

Note that the homogenization $\overline{q}_\alpha$ satisfies

$$\overline{q}_\alpha(\gamma) = \int_{l_\gamma} \alpha$$

where l_γ is the *free* geodesic loop corresponding to the *conjugacy* class of γ in $\pi_1(M)$. For, changing the basepoint p changes q_α by a bounded amount, and therefore does not change the homogenization. Then this formula is obviously true when p is chosen (for each γ) so that $L_\gamma = l_\gamma$.

A similar construction makes sense for closed manifolds M of variable negative curvature.

2.3.2. Counting quasimorphisms.

DEFINITION 2.25. Let F be a free group on a symmetric generating set S. Let w be a reduced word in S. The *big counting function* $C_w(g)$ is defined by

$$C_w(g) = \text{number of copies of } w \text{ in the reduced representative of } g$$

and the *little counting function* $c_w(\cdot)$ is defined by

$$c_w(g) = \text{max. number of } \textit{disjoint} \text{ copies of } w \text{ in the reduced representative of } g$$

A *big counting quasimorphism* is a function of the form

$$H_w(g) := C_w(g) - C_{w^{-1}}(g)$$

and a *little counting function* is a function of the form

$$h_w(g) = c_w(g) - c_{w^{-1}}(g)$$

Big counting functions were introduced by Brooks in [**27**]. We sometimes refer to C_w or H_w (and even c_w or h_w) as *Brooks functions* or *Brooks quasimorphisms.* The little counting functions, and variations on them, were introduced by Epstein–Fujiwara [**78**], who generalized them to arbitrary hyperbolic groups (although the big counting functions also generalize easily to hyperbolic groups). These two functions are related, but different, and have different advantages in different situations. We shall see that the big counting quasimorphisms are computationally simpler, and easier to deal with, whereas the little counting quasimorphisms (and their generalizations) have *uniformly* small defects, and are therefore more "powerful".

REMARK 2.26. Suppose no proper suffix of w is equal to a proper prefix. Then copies of w in any reduced word are necessarily disjoint, and $h_w = H_w$. Grigorchuk [**95**] uses the terminology "no overlapping property" to describe such words.

Every H_w and h_w is a quasimorphism. In fact, we will explicitly calculate their defects in what follows. First we must prove some preliminary statements.

LEMMA 2.27. *Let $u \in F$ be reduced. Copies of w in u are disjoint from copies of w^{-1}.*

PROOF. Suppose not, so that without loss of generality some suffix of w is equal to some prefix of w^{-1}. But in this case $w = w_1 w_2$ where $w_2 = w_2^{-1}$ which is impossible. □

Let $u \in F$ be reduced, and let $u = u_1 u_2$ as a reduced expression (i.e. there is no cancellation of the suffix of u_1 with the prefix of u_2). Say that a copy of w intersects the *juncture* of u if it overlaps both the suffix of u_1 and the prefix of u_2. By Lemma 2.27, at most one of w, w^{-1} can intersect the juncture of u.

DEFINITION 2.28. Given a reduced expression $u = u_1 u_2$ and a reduced word w, the *sign* of the expression, denoted s, is

$$s = \begin{cases} 1 & \text{if } w \text{ intersects the juncture} \\ -1 & \text{if } w^{-1} \text{ intersects the juncture} \\ 0 & \text{otherwise} \end{cases}$$

LEMMA 2.29. *Let $u = u_1 u_2$ be a reduced expression with sign s. Then*

$$h_w(u) - h_w(u_1) - h_w(u_2) = 0 \text{ or } s$$

and

$$0 \le s(H_w(u) - H_w(u_1) - H_w(u_2)) \le |w| - 1$$

PROOF. At most $|w|-1$ copies of w or w^{-1} can intersect the juncture, proving the second inequality.

To prove the first equality, for $i = 1, 2$ let U_i be a maximal disjoint configuration of copies of w in u_i. Then $U_1 \cup U_2$ is contained in $u_1 u_2$, so $c_w(u) - c_w(u_1) - c_w(u_2) \ge 0$. Conversely, let U be a maximal disjoint configuration of copies of w in $u_1 u_2$. Then either U contains one copy of w which intersects the juncture, or else it is disjoint from the juncture and decomposes as $U = U_1 \cup U_2$. Hence $c_w(u) - c_w(u_1) - c_w(u_2) \le 1$ if $s = 1$ and $c_w(u) - c_w(u_1) - c_w(u_2) \le 0$ otherwise. □

It follows that $D(H_w) \le 3(|w| - 1)$. One cannot do better than $O(|w|)$ in general, as an example like $w = abababababa$ shows. However, for little counting quasimorphisms, one obtains $D(h_w) \le 3$, and with more work one can find an even sharper estimate.

PROPOSITION 2.30. *Let w be a reduced word. Then*

(1) $D(h_w) = 0$ *if and only if* $|w| = 1$
(2) $D(h_w) = 2$ *if and only if w is of the form* $w = w_1 w_2 w_1^{-1}$, $w = w_1 w_2 w_1^{-1} w_3$ *or* $w = w_1 w_2 w_3 w_2^{-1}$ *as reduced expressions*
(3) $D(h_w) = 1$ *otherwise*

PROOF. If $|w| = 1$, the subgroup $\langle w \rangle$ generated by w is a $\mathbb{Z}$ summand of F, and h_w is just projection from F onto this summand; i.e. it is a homomorphism. Otherwise, if $w = w_1 w_2$ is a reduced expression, $h_w(w) = 1$ whereas $h_w(w_1) = h_w(w_2) = 0$. This proves the first statement.

Let $u, v \in F$ be reduced. Then we can uniquely write $u = u'x, v = x^{-1}v'$ where $u'v'$ is the reduced representative of uv. Let s_1, s_2, s_3 be the signs of the reduced expressions $u'x, x^{-1}v', u'v'$ respectively. We calculate

$$\begin{aligned} h_w(uv) - h_w(u) - h_w(v) &= h_w(uv) - h_w(u) - h_w(v) \\ &\quad - h_w(u') + h_w(u') - h_w(v') + h_w(v') + h_w(x) - h_w(x^{-1}) \\ &= (0 \text{ or } s_3) - (0 \text{ or } s_1) - (0 \text{ or } s_2) \end{aligned}$$

After possibly replacing w with w^{-1} and reversing the order of the strings, there are only nine possibilities for (s_1, s_2, s_3):

$$|h_w(uv) - h_w(u) - h_w(v)| \le \begin{cases} 0 & \text{for } (0,0,0) \\ 1 & \text{for } (1,0,0),\ (0,0,1),\ (1,-1,0),\ (1,0,1) \\ 2 & \text{for } (1,1,0),\ (1,1,1),\ (1,0,-1) \\ 3 & \text{for } (1,1,-1) \end{cases}$$

CASE $((1, 0, -1))$. If w overlaps $u'x$ and w^{-1} overlaps $u'v'$ then either some prefix of w is equal to a substring of w^{-1} or some prefix of w^{-1} is equal to a substring of w. In either case w has the form asserted by bullet (2).

CASE $((1, 1, s))$. Since w overlaps both $u'x$ and $x^{-1}v'$ we can write $w = w_1 w_2 w_3$ where either $w_2 w_3$ is the prefix of x and $w_1 w_2$ is the suffix of x^{-1} or w_3 is the prefix of x and w_1 is the suffix of x^{-1}. In the first case, $w_2^{-1} w_1^{-1}$ is the prefix of x so $w_2 = w_2^{-1}$ which is absurd. Hence we must be in the second case, and one of w_1^{-1}, w_3 is a prefix of the other.

In either case w has the form asserted by bullet (2), so we are done unless $s = -1$.

SUBCASE $((1, 1, -1))$. Without loss of generality, we can assume w is of the form $w = w_1 w_2 w_3 w_2^{-1}$ where $w_1 w_2 w_3$ is the terminal string of u' and $w_3 w_2^{-1}$ is the initial string of v'. By hypothesis, a copy of $w^{-1} = w_2 w_3^{-1} w_2^{-1} w_1^{-1}$ overlaps $y := w_1 w_2 w_3 w_3 w_2^{-1}$.

By Lemma 2.27, the subword $w_3^{-1} w_2^{-1} w_1^{-1}$ cannot overlap $w_1 w_2 w_3$ in y. Also, the subword $w_2 w_3^{-1}$ of w^{-1} cannot overlap $w_3 w_2^{-1}$ in y. Hence the w_3^{-1} in w^{-1} cannot overlap $w_1 w_2 w_3 w_3 w_2^{-1}$ at all. So if there is any overlap, either the suffix $w_2^{-1} w_1^{-1}$ of w^{-1} intersects the prefix $w_1 w_2$ of y or the prefix w_2 of w^{-1} intersects the suffix w_2^{-1} of y. But neither case can occur, again by Lemma 2.27. Hence this subcase cannot occur.

One can check that if w has the form asserted by bullet (2) then $D(h_w) \ge 2$ by example. This completes the proof. □

EXAMPLE 2.31 (monotone words).

DEFINITION 2.32. A word w is *monotone* if for each $a \in S$, at most one of a and a^{-1} appears in w.

By Proposition 2.30, for any reduced monotone word w, there is an inequality $D(h_w) \le 1$ where $D(h_w) = 1$ whenever $|w| > 1$. Notice that any reduced word of length 2 is monotone.

It is also interesting to study linear combinations of counting quasimorphisms. If w_i is a sequence of words, and t_i is a sequence of real numbers with $\sum_i |t_i| < \infty$ then $\sum_i t_i h_{w_i}$ is a quasimorphism with defect at most $2\sum_i |t_i|$. However, even if $\sum_i |t_i|$ is infinite, the function $\sum_i t_i h_{w_i}$ might still be a quasimorphism.

Definition 2.33. A family of reduced words W is *compatible* if there are words $\overline{u}, \overline{v}$ (possibly left- and right-infinite respectively) so that for each $w \in W$ there is a factorization $w = uv$ (not necessarily unique) for which each u is a suffix of $\overline{u}$ and each v is a prefix of $\overline{v}$.

Proposition 2.34. *Let $\phi = \sum_{w\in W} t(w)h_w$ for some real numbers $t(w)$. Suppose there is a finite T such that for every compatible family $V \subset W$ there is an inequality*

$$\sum_{w\in V} |t(w)| \le T$$

Then ϕ is a quasimorphism with $D(\phi) \le 3T$.

Proof. Given $u = u'x$ and $v = x^{-1}v'$, the size of $\phi(u) + \phi(v) - \phi(uv)$ can be estimated by counting copies of words $w \in W$ which overlap $u'x, x^{-1}v'$ or $u'v'$. The family of words which contribute at each overlap is a compatible family, so the claim follows. □

Example 2.35. The function

$$H := H_{aba} + H_{abba} + H_{abbba} + \cdots$$

satisfies $D(H) = 1$ (by monotonicity, and the fact that the big and small counting quasimorphisms are equal for these particular words).

Example 2.36. Let W be the family of all words in a, b (but not their inverses). There are 2^n words of length n. Define $\phi = \sum_{w\in W} 2^{-|w|}|w|^{-1}h_w$. In a compatible family, there are at most n words of length n for each n, so $D(\phi) \le 3$. On the other hand, $\sum_w 2^{-|w|}|w|^{-1} = \sum_n n^{-1} = \infty$.

Remark 2.37. Similar examples and a discussion of limits of sums of quasimorphisms are found in [**95**].

2.3.3. Rotation number.

Poincaré [**167**] introduced rotation numbers in his study of 1-dimensional dynamical systems. Let $\text{Homeo}(S^1)$ denote the group of homeomorphisms of the circle, and $\text{Homeo}^+(S^1)$ its orientation-preserving subgroup. Let G be a subgroup of $\text{Homeo}^+(S^1)$. Let $\widehat{G}$ be the preimage of G in $\text{Homeo}^+(\mathbb{R})$ under the covering projection $\mathbb{R} \to S^1$.

Note that $\widehat{G}$ is a (possibly trivial) central extension of G, and is centralized (in $\text{Homeo}^+(\mathbb{R})$) by the subgroup generated by a translation $Z : x \to x + 1$.

Definition 2.38 (Poincaré's rotation number). Given $g \in \widehat{G}$, define the *rotation number* to be

$$\text{rot}(g) = \lim_{n\to\infty} \frac{g^n(0)}{n}$$

Remark 2.39. Many authors also use the terminology "translation number" or "translation quasimorphism" for rot on $\widehat{G}$.

Rotation number is a quasimorphism:

LEMMA 2.40. rot *is a quasimorphism on* $\widehat{G}$.

PROOF. Since Z is central, $\text{rot}(Z^n a) = n + \text{rot}(a)$ for all a. Given arbitrary a, b, write $a = Z^n a', b = Z^m b'$ where $0 \le a'(0) < 1$ and $0 \le b'(0) < 1$. Of course this implies $ab = Z^{m+n} a'b'$. Then

$$0 \le \text{rot}(a') + \text{rot}(b') \le 2,\ 0 \le \text{rot}(a'b') \le 2$$

and one obtains the estimate $D(\text{rot}) \le 2$. □

In fact, one can obtain more precise information.

LEMMA 2.41. *For all* $p \in \mathbb{R}$ *and* $a, b \in \widehat{G}$ *there is an inequality*

$$p - 2 < [a, b](p) < p + 2$$

PROOF. For any p, after multiplying a, b by elements of the center if necessary (which does not change $[a, b]$) we can assume $p \le a(p), b(p) < p+1$. Then we obtain two inequalities

$$p \le a(p) \le ab(p) < a(p + 1) < p + 2$$
$$p \le b(p) \le ba(p) < b(p + 1) < p + 2$$

Let $q = ba(p)$. Then from the second inequality we obtain

$$p \le q < p + 2$$

and therefore from the first inequality,

$$q - 2 < p \le ab(p) = aba^{-1}b^{-1}(q) < p + 2 \le q + 2$$

Since p was arbitrary, so was q (up to multiplication by an element of the center). But the center commutes with $aba^{-1}b^{-1}$, so we obtain an inequality

$$q - 2 < aba^{-1}b^{-1}(q) < q + 2$$

valid for any $q \in \mathbb{R}$. This proves the Lemma. □

REMARK 2.42. Lemma 2.41 is well-known; the proof given above is essentially the same as that of Proposition 3.1 from [**197**].

It follows that there is an estimate $\text{scl}(a) \ge |\text{rot}(a)|/2$ for any $a \in \widehat{G}$. It turns out that this estimate is sharp.

THEOREM 2.43. *Let* $\text{Homeo}^+(\mathbb{R})^{\mathbb{Z}}$ *denote the full preimage of* $\text{Homeo}^+(S^1)$ *in* $\text{Homeo}^+(\mathbb{R})$. *Then* $\text{scl}(a) = |\text{rot}(a)|/2$ *in* $\text{Homeo}^+(\mathbb{R})^{\mathbb{Z}}$.

PROOF. Let b be an element which translates some elements in the positive direction and some elements in the negative direction. Then for any $p \in \mathbb{R}$ and any small $\epsilon > 0$, some conjugate of b takes p to $p + 1 - \epsilon$. Similarly, some conjugate of b^{-1} takes $b(p)$ to $b(p) + 1 - \epsilon$. It follows that for any $p \in \mathbb{R}$ and any small $\epsilon > 0$ there is a commutator which takes p to $p + 2 - 2\epsilon$.

Given a with $|\text{rot}(a)| = r$, the power a^n moves every point a distance less than $nr + 1$. It turns out that the estimate in Lemma 2.41 is sharp, in the sense that for any $p \in \mathbb{R}$ and any $|s| < 2$ one can find a commutator g such that $g(p) - p = s$. Therefore a^n can be written as a product of at most $\lfloor (nr + 1)/2 \rfloor + 1$ commutators with an element a' which fixes some point. The dynamics of a' on every complementary interval to $\text{fix}(a')$ is topologically conjugate to a translation of $\mathbb{R}$, which is the commutator of two dilations. Therefore any element a' of $\text{Homeo}^+(\mathbb{R})^{\mathbb{Z}}$ with a fixed point is a commutator. So $\mathrm{cl}(a^n) \le \lfloor (nr+1)/2 \rfloor + 2$. Dividing both sides by n, and taking the limit as $n \to \infty$ we get an inequality $\mathrm{scl}(a) \le |\mathrm{rot}(a)|/2$.

On the other hand, since a^n moves every point a distance at least $nr+1$, and by Lemma 2.41 every commutator moves every point a distance at most 2, we get an inequality $n|\mathrm{rot}(a)| \le 2 \cdot \mathrm{cl}(a^n) + 1$ and therefore $|\mathrm{rot}(a)|/2 \le \mathrm{scl}(a)$. This proves the Theorem. □

See e.g. [**70**] for more details and an extensive discussion.

Remark 2.44. Note that the group $\mathrm{Homeo}^+(S^1)$ is uniformly perfect — every element can be written as a product of at most two commutators. For, every element can be written as a product of two elements both of which have a fixed point, and (as observed in the proof of Theorem 2.43) every element of $\mathrm{Homeo}^+(S^1)$ with a fixed point is a commutator. In fact, a more detailed argument shows that every element of $\mathrm{Homeo}^+(S^1)$ is a commutator.

2.4. Bounded cohomology

2.4.1. Bar complex.

Definition 2.45. Let G be a group. The *bar complex* $C_*(G)$ is the complex generated in dimension n by n-tuples $(g_1, \ldots, g_n)$ with $g_i \in G$ and with boundary map ∂ defined by the formula

$$\partial(g_1, \ldots, g_n) = (g_2, \ldots, g_n) + \sum_{i=1}^{n-1} (-1)^i (g_1, \ldots, g_i g_{i+1}, \ldots, g_n) + (-1)^n (g_1, \ldots, g_{n-1})$$

For a coefficient group R, we let $C^*(G;R)$ denote the terms in the dual cochain complex $\mathrm{Hom}(C_*(G), R)$, and let δ denote the adjoint of ∂. The homology groups of $C^*(G;R)$ are called the *group cohomology of G with coefficients in R*, and are denoted $H^*(G;R)$.

If R is a subgroup of $\mathbb{R}$, a cochain $\alpha \in C^n(G)$ is *bounded* if

$$\sup |\alpha(g_1, \ldots, g_n)| < \infty$$

where the supremum is taken over all generators. This supremum is called the *norm* of α, and is denoted $\|\alpha\|_\infty$. The set of all bounded cochains forms a subcomplex $C_b^*(G)$ of $C^*(G)$, and its homology is the so-called *bounded cohomology* $H_b^*(G)$.

The norm $\|\cdot\|_\infty$ makes $C_b^n(G)$ into a Banach space for each n. There is a natural function on $H_b^*(G)$ defined as follows: if $[\alpha] \in H_b^*(G)$ is a cohomology class, set

$$\|[\alpha]\|_\infty = \inf \|\sigma\|_\infty$$

where the infimum is taken over all cocycles σ in the class of $[\alpha]$. If the bounded coboundaries $B_b^n(G)$ are a closed subspace of $C_b^n(G)$, this function defines a Banach norm on $H_b^n(G)$. However, it should be pointed out that $B_b^n(G)$ is *not* typically closed in $C_b^n(G)$.

There is an obvious L^1 norm on $C_*(G;\mathbb{R})$ defined in the same way as the Gromov norm for singular chains from Definition 1.11, so these chain groups may be thought of as (typically incomplete) normed vector spaces.

2.4.2. Amenable groups. Let G be a group. Recall that a *mean* on G is a linear functional on $L^\infty(G)$ which maps the constant function $f(g) = 1$ to 1, and maps non-negative functions to non-negative numbers.

DEFINITION 2.46. A group G is *amenable* if there is a G-invariant mean $\pi : L^\infty(G) \to \mathbb{R}$ where G acts on $L^\infty(G)$ by

$$g \cdot f(h) = f(g^{-1}h)$$

for all $g, h \in G$ and $f \in L^\infty(G)$.

Examples of amenable groups are finite groups, solvable groups, and Grigorchuk's groups of intermediate growth.

Bounded cohomology behaves well under amenable covers:

THEOREM 2.47 (Johnson, Trauber, Gromov). *Let*

$$1 \to H \to G \to A \to 1$$

be exact, where A is amenable. Then the natural homomorphisms $H_b^*(G;\mathbb{R}) \to H_b^*(H;\mathbb{R})^A$ are isometric isomorphisms in each dimension.

Here $H_b^*(H;\mathbb{R})^A$ denotes the A-invariant part of $H_b^*(H;\mathbb{R})$ under the action of A on H by outer automorphisms. In particular, if $H_b^*(H;\mathbb{R})$ vanishes, so does $H_b^*(G;\mathbb{R})$. We give the sketch of a proof (also see Proposition 2.65):

PROOF. Replace groups by spaces, so that X is a $K(G,1)$, and $\tilde{X}$ is a $K(H,1)$ thought of as a covering space of X with deck group A. The complex of singular bounded cochains $C_b^*(X)$ on X can be naturally identified with the complex of A-invariant singular bounded cochains $C_b^*(\tilde{X})^A$ on $\tilde{X}$. Since A is amenable, averaging over orbits defines an A-invariant projection $\pi : C_b^*(\tilde{X}) \to C_b^*(X)$. The projection π commutes with the coboundary, and is a left inverse to the pullback homomorphism defined by $\tilde{X} \to X$, and therefore the pullback homomorphism induces an isometric embedding $H_b^*(X) \to H_b^*(\tilde{X})$. The image is clearly contained in $H_b^*(\tilde{X})^A$, and in fact by averaging can be shown to coincide with it.

The proof is completed by showing that bounded group cohomology $H_b^*(G;\mathbb{R})$ is isometrically isomorphic to bounded singular cohomology $H_b^*(K(G,1);\mathbb{R})$ for any G. □

See [**117**] or [**97**] pp. 38–44 for more details.

REMARK 2.48. Theorem 2.47 is only valid for $\mathbb{R}$ coefficients, since the maps depend on averaging, which does not make sense over other coefficient groups. In particular, bounded cohomology over other coefficient groups (e.g. $\mathbb{Z}$) can be nontrivial, and even quite interesting, for some amenable groups.

An important corollary is the case that $G = A$ amenable. Since H_b^* of the trivial group is trivial, this implies that $H_b^*(A;\mathbb{R})$ vanishes identically when A is amenable.

Fibrations with amenable fiber are not so well-behaved, since spectral sequences for bounded cohomology are complicated. However, in dimension two, one has the following useful theorem of Bouarich [**19**]:

THEOREM 2.49 (Bouarich [**19**]). *Let*

$$K \to G \to H \to 1$$

be exact. Then the induced sequence on second bounded cohomology is (left) exact:

$$0 \to H_b^2(H;\mathbb{R}) \to H_b^2(G;\mathbb{R}) \to H_b^2(K;\mathbb{R})$$

In particular, if K is amenable, $H_b^2(H) \to H_b^2(G)$ is an isomorphism. We will give a proof of this theorem in § 2.7.2.

For a more detailed introduction to bounded cohomology, see Gromov's paper [**97**] or either of the references [**115**], [**157**].

2.4.3. Exact sequences and filling norms. I am grateful to Shigenori Matsumoto who provided elegant proofs of many results in this section. In the sequel, we use some of the elements of abstract functional analysis; Rudin [**180**] is a general reference.

Recall our notation $\widehat{Q}(G)$ for the vector space of all quasimorphisms on G, and $Q(G)$ for the vector subspace of homogeneous quasimorphisms. Recall that $D(\cdot)$ defines pseudo-norms on both $\widehat{Q}(G)$ and $Q(G)$ which vanish exactly on the subspace spanned by *homomorphisms* $G \to \mathbb{R}$. This subspace may be naturally identified with $H^1(G;\mathbb{R})$.

A real-valued function φ on G may be thought of as a 1-cochain, i.e. as an element of $C^1(G;\mathbb{R})$. The coboundary δ of such a function is defined by the formula

$$\delta\varphi(a,b) = \varphi(a) + \varphi(b) - \varphi(ab)$$

At the level of norms, there is an equality, $\|\delta\varphi\|_\infty = D(\varphi)$. It follows that the coboundary of a quasimorphism is a *bounded* 2-cocycle.

THEOREM 2.50 (Exact sequence). *There is an exact sequence*

$$0 \to H^1(G;\mathbb{R}) \to Q(G) \to H_b^2(G;\mathbb{R}) \to H^2(G;\mathbb{R})$$

PROOF. There is an exact sequence of chain complexes

$$0 \to C_b^* \to C^* \to C^*/C_b^* \to 0$$

and an associated long exact sequence of cohomology groups. A bounded homomorphism to $\mathbb{R}$ is trivial, hence $H_b^1(G;\mathbb{R}) = 0$ for any group G. A function φ on G is in $\widehat{Q}(G)$ if and only if $\delta\varphi$ is in C_b^2. Moreover, any two quasimorphisms which differ by a bounded amount have the same homogenization. Hence

$$H^1(C^*/C_b^*) = \widehat{Q}/C_b^1 \cong Q$$

□

EXAMPLE 2.51. Recall from § 2.3.3 that rot is a homogeneous quasimorphism on the group $\text{Homeo}^+(\mathbb{R})^{\mathbb{Z}}$, which is our notation for the group of homeomorphisms of $\mathbb{R}$ which are periodic with period 1. Further recall that this group is the universal central extension of $\text{Homeo}^+(S^1)$. The function rot does not descend to a well-defined real-valued function on $\text{Homeo}^+(S^1)$, but it is well-defined mod $\mathbb{Z}$. However, the coboundary $[\delta\text{rot}]$, as a class in $H_b^2(\text{Homeo}^+(\mathbb{R})^{\mathbb{Z}})$, can be pulled back from a class in $H_b^2(\text{Homeo}^+(S^1))$. By Theorem 2.50, the image of this class in $H^2(\text{Homeo}^+(S^1))$ is a nontrivial class, called the *Euler class*. The L^∞ norm of this class is $1/2$ (compare with Theorem 2.43). This fact is otherwise known as the *Milnor–Wood inequality* ([**154**],[**204**]), and is usually stated in the following way:

THEOREM 2.52 (Milnor–Wood inequality). *Let S be a closed, oriented surface of genus g, and let $\rho : \pi_1(S) \to \text{Homeo}^+(S^1)$ be an action of $\pi_1(S)$ on a circle by homeomorphisms. Let $[e] \in H^2(S)$ be the pullback of the generator of $H^2(\text{Homeo}^+(S^1);\mathbb{Z})$. Then there is an inequality*

$$|[e](S)| \le -\chi^-(S)$$

For ease of notation, we abbreviate $C_*(G;\mathbb{R})$ in what follows by C_*. Similarly, denote cycles and boundaries with real coefficients by Z_* and B_* respectively. Then

$$0 \to Z_2 \to C_2 \to B_1 \to 0$$

is exact. Since C_2 is normed, and Z_2 is a normed subspace, B_1 inherits a quotient norm.

Observe that if $a \in [G,G]$ then $a \in B_1$ when thought of as a generator of C_1. For example, if $a = [x,y]$ then

$$\partial((xyx^{-1},x) + ([x,y],y) - (x,y)) = [x,y]$$

In general, a one-vertex triangulation of a surface of genus g with one boundary component exhibits a product of g commutators as an element of B_1.

DEFINITION 2.53. Let $a \in B_1(G;\mathbb{R})$. The *Gersten boundary norm* (or just the *Gersten norm* or the *boundary norm*) of a, denoted $\|a\|_B$, is defined by

$$\|a\|_B = \inf_{\partial A = a} \|A\|_1$$

where the infimum is taken over all 2-chains $A \in C_2(G;\mathbb{R})$ with boundary a, and $\|A\|_1$ denotes the usual L^1 norm.

REMARK 2.54. Gersten calls his norm a *filling norm* in [**90**]. However, we reserve this name for a suitable homogenization of $\|\cdot\|_B$.

It is important to note that this quotient is really a norm and not just a pseudonorm, since ∂ is a bounded operator on C_2 of norm 3, and therefore $\|a\|_1 \le 3\|a\|_B$. In particular, Z_2 is closed in C_2 in the L^1 norm.

REMARK 2.55. We can define $C_*^{l_1}$ to be the completion of C_* with respect to the L^1 norm. The boundary map ∂ extends continuously to $C_*^{l_1}$, and we let $Z_*^{l_1}$ and $B_*^{l_1}$ denote the kernel and image of ∂ respectively. The exact sequence

$$0 \to Z_2^{l_1} \to C_2^{l_1} \to B_1^{l_1} \to 0$$

defines a quotient norm on $B_1^{l_1}$ and thereby on B_1 under inclusion $B_1 \to B_1^{l_1}$. However, in general there is a *strict* inclusion $\overline{Z}_2 \subset Z_2^{l_1}$, where $\overline{Z}_*$ denotes the completion of Z_* in the L^1 norm, and therefore the norm B_1 inherits as a subspace of $C_2^{l_1}/Z_2^{l_1}$ will be typically *smaller* than $\|\cdot\|_B$.

In fact there is an important special case in which the two norms on B_1 are the same. Matsumoto–Morita [**150**] say that the chain complex C_* satisfies condition 1-UBC if there is a positive constant $K > 0$ such that $K\|a\|_B \le \|a\|_1$ for all $a \in B_1$. Note that this is equivalent to the condition that the norms $\|\cdot\|_1$ and $\|\cdot\|_B$ induce the same topology on B_1. Under this circumstance, there is an equality $\overline{Z}_2 = Z_2^{l_1}$. In fact, Theorem 2.8 from [**150**] implies that condition 1-UBC is equivalent to injectivity of the map $H_b^2 \to H^2$. By Theorem 2.50, this is equivalent to $Q(G)/H^1(G) = 0$, a situation which is largely orthogonal to the focus of this book.

We now identify the dual space of B_1 with respect to the norm $\|\cdot\|_B$.

LEMMA 2.56. *The dual of B_1 with respect to the $\|\cdot\|_B$ norm is $\widehat{Q}(G)/H^1(G;\mathbb{R})$, and the operator norm on the dual is equal to $D(\cdot) = \|\delta \cdot\|_\infty$.*

PROOF. In the sequel, if V is a normed vector space, we denote the space of bounded linear functionals on V with the operator norm by V'.

By definition of the quotient norm, an element f of B_1' determines $F \in C_2'$ with the same operator norm, vanishing on Z_2, by the formula $F(A) = f(\partial A)$. Since F vanishes on Z_2, it is a coboundary; hence $F = \delta\phi$ where $\phi \in C^1$ is unique up to an element of H^1. Since F is bounded, ϕ is a quasimorphism, and we have defined $B_1' \to \widehat{Q}/H^1$ (note that the restriction of ϕ to B_1 is equal to f). This map is evidently injective and surjective, and is therefore an isomorphism of vector spaces.

It remains to identify the norm. Let $b \in B_1$ be an element with $\|b\|_B = 1$, so there is $A \in C_2$ with $\partial A = b$ and $\|A\|_1 - 1 < \epsilon$. Express A as $A = \sum_j r_j(g_j, h_j)$ with $r_j \in \mathbb{R}$, and $\sum_j |r_j| - 1 < \epsilon$. By the triangle inequality,

$$\begin{aligned} |F(A)|/(1+\epsilon) \le \sup_j |F(g_j,h_j)| = \sup_j |\delta\phi(g_j,h_j)| = \sup_j |\phi(\partial(g_j,h_j))| \\ = \sup_j |\phi(g_jh_j) - \phi(g_j) - \phi(h_j)| \le D(\phi) \end{aligned}$$

so we deduce that the operator norm of F (and therefore that of f) is $\le D(\phi)$.

Conversely, let $g_1, g_2 \in G$ be arbitrary. Then (except in degenerate cases) $\partial(g_1,g_2) = g_1 + g_2 - g_1g_2$ has L^1 norm equal to 3, and therefore

$$1 \ge \|\partial(g_1,g_2)\|_B \ge \frac{1}{3}\|\partial(g_1,g_2)\|_1 = 1$$

But $F(g_1,g_2) = \phi(g_1) + \phi(g_2) - \phi(g_1g_2)$, so by the definition of the defect there are $g_1, g_2 \in G$ with $\|\partial(g_1,g_2)\|_B = 1$ for which $|F(g_1,g_2)|$ is arbitrarily close to $D(\phi)$. This implies that the operator norm of F is at least equal to $D(\phi)$, and together with the previous inequality, this shows that the operator norm of F is exactly equal to the defect of ϕ, as claimed. □

We deduce the following corollary:

COROLLARY 2.57. *The space $\widehat{Q}/H^1$ with its defect norm is a Banach space, and is isometric to the dual of $C_2^{l_1}/\overline{Z}_2$ with its L^1 norm.*

PROOF. By Lemma 2.56, we know that $\widehat{Q}/H^1$ with its defect norm is the dual of B_1 with its $\|\cdot\|_B$ norm, which by definition is equal to the dual of C_2/Z_2 with its L_1 norm. If X is a normed vector space, and Y is a closed normed vector subspace, the dual $(X/Y)'$ is isometrically isomorphic to the dual $(\overline{X}/\overline{Y})'$ where the overline denotes completion with respect to the norm. In our case, $C_2^{l_1}$ and $\overline{Z}_2$ are the completions of C_2 and Z_2 in the L^1 norm, so the second claim of the corollary follows.

The dual space of a normed vector space is always a Banach space. Hence the first claim follows already from Lemma 2.56. □

Since homogeneity is a closed condition, the quotient Q/H^1 is a Banach subspace of $\widehat{Q}/H^1$. We refer to the Banach topology on this space as the *defect topology*. *A priori*, there is a natural pseudo-norm on H_b^2. We will see shortly that this pseudo-norm is actually a norm (this fact is due to Matsumoto–Morita [**150**]). Theorem 2.50 shows that δ is an injection of Q/H^1 into H_b^2. The next lemma describes how the norm behaves under δ :

LEMMA 2.58. *Let $\phi \in Q(G)$. Then*

$$D(\phi) \geq \|[\delta\phi]\|_\infty \geq \frac{1}{2}D(\phi)$$

PROOF. By definition, $\|[\delta\phi]\|$ is the infimum of the L^∞ norm of all bounded 2-cocycles A which are cohomologous to $\delta\phi$. Now any such A is of the form δf for some unique (not necessarily homogeneous) quasimorphism f for which $f-\phi \in C^1_b$. In particular, ϕ is the homogenization of f, and we have an inequality

$$\|[\delta\phi]\|_\infty = \inf_{f-\phi\in C^1_b} D(f) \leq D(\phi)$$

Since any quasimorphism can be antisymmetrized without increasing its defect, it suffices to take the infimum over antisymmetric f.

Let $a,b \in G$ be such that $|\delta\phi(a,b)|$ is very close to $D(\phi)$. Recall from the proof of Lemma 2.24 that $a^{2n}b^{2n}(ab)^{-2n}$ can be written as a product of at most n commutators. Since f is antisymmetric, it follows that $|f(a^{2n}b^{2n}(ab)^{-2n})| \leq (4n-1)D(f)$. Since $f - \phi \in C^1_b$, there is a constant C, independent of a,b and n, so that $|f(a^{2n}b^{2n}(ab)^{-2n}) - \phi(a^{2n}b^{2n}(ab)^{-2n})| \leq C$. Moreover, by homogeneity, $|\phi(a^{2n}b^{2n}(ab)^{-2n}) - 2n\delta\phi(a,b)| \leq 2D(\phi)$ and therefore

$$\lim_{n\to\infty} \frac{|\phi(a^{2n}b^{2n}(ab)^{-2n})|}{2n} = |\delta\phi(a,b)|$$

which is arbitrarily close to $D(\phi)$. Putting this together, we get an estimate

$$D(\phi) \leq 2D(f)$$

and the lemma is proved. □

It is convenient to explicitly record the following corollary:

COROLLARY 2.59. *Let $f \in \widehat{Q}(G)$ with homogenization $\phi \in Q(G)$. Then*

$$D(f) = \|\delta f\|_\infty \geq \|[\delta\phi]\|_\infty \geq \frac{1}{2}D(\phi)$$

REMARK 2.60. Lemma 2.58 and its Corollary can be restated in homological language. The following argument is due to Shigenori Matsumoto. Since $C^1_b \cap H^1 = 0$, we can think of C^1_b as a subspace of $\widehat{Q}/H^1$. We have already shown in Corollary 2.57 that $\widehat{Q}/H^1$ can be identified with the dual $(C^{l_1}_2/\overline{Z}_2)'$. What is the image $\delta(C^1_b)$ in this dual space? First we make an observation.

LEMMA 2.61. *The boundary map $\partial : C^{l_1}_2 \to C^{l_1}_1$ has a (bounded) cross-section σ defined by the formula*

$$\sigma(g) = \frac{1}{2}(g,g) + \frac{1}{4}(g^2,g^2) + \cdots$$

PROOF. The proof is immediate. □

From this it follows that $B^{l_1}_1 = C^{l_1}_1$ as abstract vector spaces. Moreover, Lemma 2.61 shows that $\|b\|_B \leq \|b\|_1$ for $b \in C^{l_1}_1$. Since we also have $\|b\|_B \geq \frac{1}{3}\|b\|_1$, this shows that the quotient norm and the L^1 norm on $C^{l_1}_1$ are equivalent (though not necessarily isometric).

The dual of $C^{l_1}_1$ with its L^1 norm is C^1_b with its L^∞ norm. Dualizing $Z^{l_1}_2 \to C^{l_1}_2 \to C^{l_1}_1$ shows that the image $\delta(C^1_b)$ is equal to $(C^{l_1}_2/Z^{l_1}_2)'$. Since $\widehat{Q}/H^1 = (C^{l_1}_2/\overline{Z}_2)'$, if we give $\widehat{Q}/(C^1_b \oplus H^1) = (\widehat{Q}/H^1)/C^1_b$ its quotient norm, we obtain an isometric isomorphism

$$\widehat{Q}/(C^1_b \oplus H^1) \xrightarrow{\delta} (Z^{l_1}_2/\overline{Z}_2)'$$

As vector spaces, Q/H^1 and $\widehat{Q}/(C_b^1 \oplus H^1)$ are naturally isomorphic; in this language, Lemma 2.58 says that their norms differ at most by a factor of 2.

Unfortunately, the Banach space $Q(G)/H^1(G;\mathbb{R})$ is typically very big, even if G is finitely presented. We give some examples to illustrate this phenomenon for the case that G is free.

EXAMPLE 2.62 (Free group). Let F denote the free group on two generators a, b. Let $w_n = ab^na$ for each positive integer n. For each $f : \mathbb{N} \to \{0,1\}$ define

$$\overline{H}_f = \sum_n f(n)\overline{H}_{w_n}$$

where each H_{w_n} is the big counting function (see Definition 2.25), and the overline denotes homogenization. Since the words are not nested, $D(H_f) = 1$ for each f (compare with Example 2.35), and therefore $D(\overline{H}_f) \le 2$ by Corollary 2.59. If $f \ne g$ then if n is in the support of f but not g (say), we have

$$(\overline{H}_f - \overline{H}_g)(ab^na) = 1$$

so the difference is nontrivial. On the other hand, since $\overline{H}_f$ and $\overline{H}_g$ vanish on both a and b, they are not in H^1. It follows that $D(\overline{H}_f - \overline{H}_g)$ is positive, and since they are both integer valued, the defect is at least 1. In other words, we have constructed a subset of $Q(F)/H^1(F)$ of cardinality $2^{\aleph_0}$ which is discrete in the defect topology. In particular, Q/H^1 is not separable.

EXAMPLE 2.63 (Density). Jason Manning constructed an explicit example of a vector in $Q(F)/H^1(F)$ which is not in the closure (in the defect topology) of the span of Brooks quasimorphisms. For each n let $w_n = [a^nb^na^{-n}, b^{-n}]$. Then $\overline{H}_v(w_n) = \overline{h}_v(w_n) = 0$ where $\overline{H}_v$ and $\overline{h}_v$ denote the homogenizations of the big and small counting functions, whenever v is a word of length $\le n$. Now, define

$$\overline{H} = \sum_i \overline{H}_{w_i}$$

Since the w_i and their inverses do not overlap, one can estimate $D(\overline{H}) \le 6$. Now suppose $\overline{H}'$ is a finite linear combination of homogenized counting quasimorphisms (of either sort). Then there is an n such that $\overline{H}'(w_n) = 0$ but $\overline{H}(w_n) = 1$. Since each w_n is a commutator, by Lemma 2.24 it follows that $D(\overline{H}' - \overline{H}) \ge 1$.

EXAMPLE 2.64 (Pullbacks). Let $F_3 = \langle a,b,c\rangle$ and $F_2 = \langle a,b\rangle$. Let $p : F_3 \to F_2$ be the obvious retraction, obtained by killing c. Let $h \in Q(F_2)$ be the homogenization of the Brooks function h_{ab}, and let $p^*h \in Q(F_3)$ denote the pullback. Then p^*h is not in the closure of the span of Brooks quasimorphisms. To see why, consider the elements $w_n := a^nca^{-n}b^{-1}a^nc^{-1}a^{-n}b$ and $w_n' := a^{n-1}ca^{-n}b^{-1}a^nc^{-1}a^{1-n}b$. The element w_n is in the kernel of p, but $p(w_n') = a^{-1}b^{-1}ab$ so $p^*h_{ab}(w_n) = 0$ whereas $p^*h_{ab}(w_n') = 1$. Note further that each w_n is a commutator, and each w_n' is a product of two commutators, and therefore satisfies $\mathrm{scl}(w_n') \le 3/2$. Notice that for any word v we must have $h_v(w_n) = h_v(w_n')$ for sufficiently large n (and similarly for H_v). It follows that p^*h cannot be approximated in defect by the homogenization of a finite linear combination of Brooks quasimorphisms (of either kind). This example is obviously not sporadic; a similar argument shows that if $p : F \to G$ is surjective with nontrivial kernel, and $h \in Q(G)$ is not in $H^1(G)$, then p^*h is never in the closure of the span of Brooks quasimorphisms.

If G is amenable, Theorem 2.47 shows that $H_b^2(G;\mathbb{R}) = 0$ and therefore $Q(G) = H^1(G;\mathbb{R})$; in other words, every homogeneous quasimorphism on an amenable group is a homomorphism to $\mathbb{R}$. For completeness, we give a self-contained proof of this fact.

Proposition 2.65. *Let G be amenable. Then every homogeneous quasimorphism on G is a homomorphism to $\mathbb{R}$.*

Proof. Let $\phi : G \to \mathbb{R}$ be a quasimorphism. We will construct a homomorphism which differs from ϕ by a bounded amount; this is enough to prove the proposition. Let $\mathbb{R}^{G\times G}$ be the space of real valued functions on $G \times G$, with the topology of pointwise convergence. A function $\phi : G \to \mathbb{R}$ determines an element $\Phi : G \times G \to \mathbb{R}$ by the formula

$$\Phi(a, b) = \phi(a) - \phi(b)$$

The group G acts on $G \times G$ diagonally: $g(a, b) = (ga, gb)$ and thus on $\mathbb{R}^{G\times G}$. For any $g \in G$, we have $g\Phi(a, b) = \phi(ga) - \phi(gb)$ and therefore

$$|g\Phi(a, b) - \Phi(a, b)| \le 2D(\phi)$$

Hence the convex hull of the orbit $G\Phi$ is a compact, convex, G-invariant subset of $\mathbb{R}^{G\times G}$. Note that Φ has the property that $\Phi(a, b) + \Phi(b, c) = \Phi(a, c)$ for any $a, b, c \in G$. In particular, Φ vanishes on any (a, a) and is antisymmetric in its arguments. This property is invariant under the action of G, and preserved under linear combinations and limits, and therefore holds for any element of the closed convex hull of $G\Phi$. This part of the argument does not use the fact that G is amenable.

If G is amenable, any linear action by G on a topological vector space which leaves invariant a compact, convex subset must have a global fixed point in that set; basically, the barycenter of any bounded orbit, weighted by the invariant mean, is G-invariant. If Ψ is such a G-invariant function we can define $\psi : G \to \mathbb{R}$ by $\psi(a) = \Psi(a, \mathrm{id})$. Since Ψ is G-invariant, $\psi(ab) = \Psi(ab, \mathrm{id}) = \Psi(b, a^{-1})$. But $\Psi(b, a^{-1}) + \Psi(a^{-1}, \mathrm{id}) = \Psi(b, \mathrm{id})$ so $\psi(ab) = \psi(b) - \psi(a^{-1})$. Since

$$\psi(a^{-1}) = \Psi(a^{-1}, \mathrm{id}) = \Psi(\mathrm{id}, a) = -\Psi(a, \mathrm{id}) = -\psi(a)$$

we are done. □

2.4.4. Antisymmetrization and orientations. In singular homology, simplices are *marked* by a total ordering of the vertices. Similarly, in group homology, generators of the bar complex are *ordered* tuples of group elements. Given a simplex Δ^n, the symmetric group S_{n+1} acts on Δ^n by permuting the vertices. There is a chain map $s : C_* \to C_* \otimes \mathbb{Q}$ defined on a generator σ of C_n by

$$s(\sigma) = \frac{1}{(n+1)!} \sum_{g \in S_{n+1}} \mathrm{sign}(g)\sigma \circ g$$

where $\mathrm{sign}(g)$ is ± 1 depending on whether $g : \Delta^n \to \Delta^n$ is orientation preserving or reversing. We can define a similar chain map from the bar complex $C_*(G) \otimes \mathbb{Q}$ to itself.

The chain map s is chain homotopic to id, and therefore induces an isomorphism in homology over $\mathbb{Q}$ or $\mathbb{R}$. Moreover, this chain map has operator norm 1 in each dimension with respect to the L^1 norm.

In dimension 1, the map s replaces an element $a \in G$ with the sum

$$s : a \to \frac{1}{2}(a - a^{-1})$$

It follows that if f' is the antisymmetrization of a 1-cochain f, there is an equality

$$f'(a) = f(s(a))$$

that is, $f' = s^* f$ where s^* is the adjoint of s in dimension 1. The observation in § 2.2.2 that antisymmetrization of quasimorphisms does not increase defect is dual to the the observation that s has operator norm 1.

This discussion is most relevant when one considers bounded cohomology over other coefficient groups, for instance over $\mathbb{Z}$. One can neither (anti)symmetrize chains nor cochains over $\mathbb{Z}$, and therefore some of the estimates we obtain in this section are no longer valid in greater generality.

2.5. Bavard's Duality Theorem

2.5.1. Banach duality and filling norms. In the last section, we defined the Gersten boundary norm, and identified its dual space. By an application of the Hahn–Banach Theorem, Lemma 2.56 lets us reinterpret the Gersten boundary norm in terms of quasimorphisms.

Corollary 2.66. *Let $a \in [G, G]$ so that $a \in B_1$ as a cycle. Then*

$$\|a\|_B = \sup_{\phi \in \widehat{Q}(G)/H^1(G;\mathbb{R})} \frac{|\phi(a)|}{D(\phi)}$$

To relate the Gersten norm to stable commutator length, we must homogenize.

Definition 2.67. Define the *filling norm*, denoted $\text{fill}(a)$ to be the homogenization of $\|a\|_B$. That is,

$$\text{fill}(a) = \lim_{n\to\infty} \frac{\|a^n\|_B}{n}$$

Remark 2.68. Some authors refer to $\text{fill}(\cdot)$ as the *stable* filling norm, to distinguish it from the Gersten filling norm.

It is not quite true that the function $\|a^n\|_B$ is subadditive in n. However, for any r, s there is an identity $\partial(a^r, a^s) = a^r + a^s - a^{r+s}$ and therefore $\|a^{r+s}\|_B \leq \|a^r\|_B + \|a^s\|_B + 1$. This is enough to show that the limit exists in Definition 2.67.

Using the estimates proved in Chapter 1, we can relate scl and $\text{fill}(\cdot)$ in a straightforward manner:

Lemma 2.69 (Bavard, Prop. 3.2. [**8**]). *There is an equality*

$$\text{scl}(a) = \frac{1}{4}\text{fill}(a)$$

Proof. An expression of a^n as a product of commutators lets us construct an orientable surface S with one boundary component, and a homomorphism $\varphi : \pi_1(S) \to G$ with $\varphi_* \partial S = a^n$ in π_1. We can find a triangulation of S with $4 \cdot \text{genus}(S) - 1$ triangles, where one edge maps to the boundary, and therefore

$$\|a^n\|_B \leq 4 \cdot \text{cl}(a^n) - 1$$

Dividing both sides by n, and taking the limit as $n \to \infty$ gives the inequality

$$\text{fill}(a) \leq 4 \cdot \text{scl}(a)$$

Conversely, let A be a chain with $\partial A = a$ with $\|A\|_1$ close to $\|a\|_B$. Let V be the finite dimensional subspace of $C_2(G;\mathbb{R})$ consisting of 2-chains with support contained in the support of A. Since V is a rational subspace, and a is a rational chain, the subspace $V \cap \partial^{-1}(a)$ contains rational points arbitrarily close to A (compare with Remark 1.5). So we may assume A is rational, after changing its norm an arbitrarily small amount. After scaling by some integer, we may assume A is an integral chain with $\partial A = na$ for which the ratio $\|A\|_1/n\|a\|_B$ is very close to 1.

As in Example 1.4, there is an orientable surface S and a chain A_S representing the fundamental class of S, and a map $\varphi : \pi_1(S) \to G$ sending boundary components to powers of conjugates of a, and such that $\varphi_*(A_S) = A$. Moreover, by construction, $\|A_S\|_1 = \|A\|_1$.

By Theorem 1.14 and Lemma 2.10 we have an inequality

$$\frac{\|A_S\|_1}{n} \geq \frac{-2\chi(S)}{n} \geq 4 \cdot \mathrm{scl}(a)$$

But $\|A_S\|_1/n$ may be taken to be arbitrarily close to $\|a\|_B$. Homogenizing the left hand side (and using the fact that the right hand side is homogeneous by definition) we obtain

$$\mathrm{fill}(a) \geq 4 \cdot \mathrm{scl}(a)$$

Putting this together with the earlier inequality, we are done. □

2.5.2. Bavard's Duality Theorem. We are now in a position to relate quasimorphisms and stable commutator length by means of Bavard's Duality Theorem:

THEOREM 2.70 (Bavard's Duality Theorem, [**8**]). *Let G be a group. Then for any $a \in [G,G]$, we have an equality*

$$\mathrm{scl}(a) = \frac{1}{2} \sup_{\phi \in Q(G)/H^1(G;\mathbb{R})} \frac{|\phi(a)|}{D(\phi)}$$

PROOF. For the sake of legibility, we suppress G in our notation in what follows. By Corollary 2.66 there is a duality

$$\|a\|_B = \sup_{\phi \in \widehat{Q}/H^1} \frac{|\phi(a)|}{D(\phi)}$$

Homogenizing and applying Lemma 2.69, we obtain an equality

$$\mathrm{scl}(a) = \frac{1}{4} \lim_{n\to\infty} \left(\sup_{\phi \in \widehat{Q}/H^1} \frac{|\phi(a^n)|}{nD(\phi)} \right)$$

Recall that in Lemma 2.21 we obtained the estimate $|\phi(a^n) - \overline{\phi}(a^n)| \leq D(\phi)$ where $\overline{\phi}$ denotes the homogenization of ϕ. It follows that for each n and any $\phi \in \widehat{Q}$ there is an inequality

$$\frac{|\phi(a^n) - \overline{\phi}(a^n)|}{nD(\phi)} \leq n^{-1}$$

Parsing this, for each n let ϕ_{n_i} be a sequence of elements in $\widehat{Q}(G)$ such that $\phi_{n_m}(a^n)/nD(\phi_{n_m})$ is within m^{-1} of the supremum. Then $\overline{\phi}_{n_m}(a^n)/nD(\phi_{n_m})$ is within $m^{-1} + n^{-1}$ of the supremum. Using $\overline{\phi}(a^n)/n = \overline{\phi}(a)$ and passing to a diagonal subsequence, we obtain

$$\mathrm{scl}(a) = \frac{1}{4} \sup_{\phi \in \widehat{Q}/H^1} \frac{|\overline{\phi}(a)|}{D(\phi)}$$

By Corollary 2.59, we get an inequality

$$\mathrm{scl}(a) \le \frac{1}{2} \sup_{\phi \in Q/H^1} \frac{|\phi(a)|}{D(\phi)}$$

On the other hand, for any homogeneous quasimorphism ϕ, if a^n is a product of m commutators then

$$|\phi(a^n)| \le 2mD(\phi)$$

so we get an inequality in the other direction, and the theorem is proved. □

2.6. Stable commutator length as a norm

In this section we show that scl can be extended in a natural way to a pseudo-norm on (a suitable quotient of) B_1. Moreover Bavard duality holds more generally in this broader context, thus revealing it as a genuine duality theorem (in the usual sense of functional analysis).

2.6.1. Definition.

DEFINITION 2.71. Let G be a group, and $a_i \in G$ for $1 \le i \le m$ a finite collection of elements. If the product of the a_i is in $[G, G]$, then define $\mathrm{cl}(a_1 + a_2 + \cdots + a_m)$ to be the smallest number of commutators whose product is equal to an expression of the form

$$a_1 t_1 a_2 t_1^{-1} \cdots t_{m-1} a_m t_{m-1}^{-1}$$

for some elements $t_i \in G$. Then define

$$\mathrm{scl}(\sum_i a_i) = \lim_{n \to \infty} \frac{\mathrm{cl}(\sum_i a_i^n)}{n}$$

Geometrically, if $\pi_1(X) = G$, and γ_i is a loop in X representing the conjugacy class of a_i, then $\mathrm{cl}(\sum_i a_i)$ is the least genus of a surface with m boundary components which maps to X in such a way that the ith boundary component wraps once around γ_i.

REMARK 2.72. If the product of the a_i has order n in $H_1(G; \mathbb{Z})$, define $\mathrm{scl}(\sum a_i) = \frac{1}{n}\mathrm{scl}(\sum a_i^n)$, and otherwise define $\mathrm{scl}(\sum a_i) = \infty$.

In fact, it it not immediately obvious that the limit in Definition 2.71 exists, since the function $\mathrm{cl}_n(\sum a_i) := \mathrm{cl}(\sum a_i^n)$ is not subadditive as a function of n. We address this issue in the next lemma.

LEMMA 2.73. *The limit in Definition 2.71 exists when it is defined (i.e. when the product of the a_i are in $[G, G]$).*

PROOF. If $\sum a_i$ has m terms, define $\mathrm{cl}_{n,m} = \mathrm{cl}(\sum a_i^n) + (m-1)$. Then (for fixed m) the function $\mathrm{cl}_{n,m}$ *is* subadditive as a function of n. For, if S_{n_1}, S_{n_2} are surfaces with m boundary components, each of which wraps n_1 and n_2 times respectively around each of m loops, then they can be tubed together by adding m rectangles to produce a surface S' with m boundary components, each of which wraps $n_1 + n_2$ times around each of the m loops, and satisfies $\mathrm{genus}(S') = \mathrm{genus}(S_{n_1}) + \mathrm{genus}(S_{n_2}) + (m-1)$. On the other hand, for fixed m, there is an equality

$$\lim_{n \to \infty} \frac{\mathrm{cl}(\sum a_i^n)}{n} = \lim_{n \to \infty} \frac{\mathrm{cl}(\sum a_i^n) + (m-1)}{n}$$

the right hand limit exists by the subadditivity of $\mathrm{cl}_{n,m}$, and therefore the left hand side does too. □

Given a space X and loops $\gamma_i : S^1 \to X$ we say that a map $f : S \to X$ is *admissible* if there is a commutative diagram:

$$\begin{array}{ccc} \partial S & \xrightarrow{\quad i \quad} & S \\ \partial f \downarrow & & f \downarrow \\ \coprod_i S^1 & \xrightarrow[\coprod \gamma_i]{} & X \end{array}$$

for which there is an integer $n(S)$ such that

$$\partial f_*[\partial S] = n(S)[\coprod_i S^1]$$

(note that the existence of an integer $n(S)$ is not automatic from the commutativity of the diagram, when there is more than one γ_i).

One has the following analogue of Proposition 2.10.

PROPOSITION 2.74. *Let $\pi_1(X) = G$, and for $1 \le i \le m$, let $\gamma_i : S^1 \to X$ be a loop representing the conjugacy class of $a_i \in G$. Then*

$$\mathrm{scl}(\sum_i a_i) = \inf_S \frac{-\chi^-(S)}{2n(S)}$$

where the infimum is taken over all admissible maps as above.

PROOF. The proof is almost identical to that of Proposition 2.10. An inequality in one direction follows from the definition, at least if one uses the "corrected" function $\mathrm{cl}_{n,m}$ in place of cl_n (see Lemma 2.73). To obtain the inequality in the other direction, let $f : S \to X$ be an admissible map of a surface. Without loss of generality, one may restrict attention to the case that each component of S_i has at least one boundary component mapping with nontrivial degree to some γ_i. Fix some big (even) integer N, and construct connected covers T_i of each S_i of degree $2N$, each with at most twice as many boundary components as S_i. The T_i may be surgered to have exactly m boundary components, each mapping to some γ_i with degree $2Nn(S)$ by gluing on only a constant number of rectangles, and thereby raising $-\chi$ by an amount which is independent of N. The reverse inequality follows. $\square$

A surface realizing the infimum in Proposition 2.74 is called *extremal* (compare with Definition 2.11).

From the geometric perspective it is clear that $\mathrm{scl}(\sum a_i)$ depends only on the conjugacy class of each term a_i, and is commutative in its arguments.

LEMMA 2.75. scl *satisfies the identity*

$$\mathrm{scl}(a^n + \sum a_i) = \mathrm{scl}(\underbrace{a + \cdots + a}_{n} + \sum a_i)$$

for any non-negative integer n and any $a, a_i \in G$.

PROOF. We use Proposition 2.74. Let X be a space with $\pi_1(X) = G$ and let γ be a loop representing the conjugacy class of a. Let S be a surface mapping to X, with n boundary components each wrapping around γ a total of m times, for some large m, and the rest wrapping around loops γ_i corresponding to the conjugacy classes of the a_i. The distinct boundary components wrapping around γ can be tubed together at the cost of raising $-\chi^-(S)$ by $n-1$, which can be taken to be arbitrarily small compared to m. This establishes an inequality in one direction.

Conversely, if S is a surface mapping to X with one boundary component wrapping some number of times around γ^n and the rest around the γ_i, take n copies of S to obtain the inequality in the other direction. □

Similarly we have the following.

LEMMA 2.76. scl *satisfies the identity*

$$\text{scl}(a + a^{-1} + \sum a_i) = \text{scl}(\sum a_i)$$

for any $a, a_i \in G$.

PROOF. Let X, γ, γ_i be as before. Let S be a surface whose boundary wraps around the various γ_i. Let A be an annulus from γ to γ^{-1} and let S' be the disjoint union of S with some number of parallel copies of A. Then $-\chi^-(S) = -\chi^-(S')$.

Conversely, suppose S is a surface with one boundary component ∂_1 bounding γ^m and one component ∂_2 bounding γ^{-m}. Glue ∂_1 to ∂_2 to obtain a surface S' with $-\chi^-(S) = -\chi^-(S')$. □

By abuse of notation we define $\text{scl}(\sum_i a_i - a) := \text{scl}(\sum_i a_i + a^{-1})$. It follows from Lemma 2.75 and Lemma 2.76 that for any a, a_i and for any equality $n = \sum_i n_i$ over $\mathbb{Z}$ there is a corresponding equality

$$\text{scl}(a^n + \sum_j a_j) = \text{scl}(\sum_i a^{n_i} + \sum_j a_j)$$

Moreover, for any integer n, there is an equality

$$|n| \, \text{scl}(\sum a_i) = \text{scl}(\sum na_i) = \text{scl}(\sum a_i^n)$$

Consequently scl can be extended by linearity on rays to rational chains $\sum_i r_i a_i$ representing 0 in $H_1(G;\mathbb{Q})$. Since scl is subadditive on rational chains, it extends continuously in a unique way to a pseudo-norm on the real vector space $B_1(G)$.

Recall from § 2.5.1 that we defined the Gersten norm $\|\cdot\|_B$ on B_1 by the equality

$$\|a\|_B = \inf_{\partial A = a} \|A\|_1$$

where $a \in B_1$ and $A \in C_2$. Then for an element $g \in [G, G]$ we defined the (stable) filling norm by the formula

$$\text{fill}(g) = \lim_{n\to\infty} \frac{\|g^n\|_B}{n}$$

One can extend fill to all of B_1. First extend fill to integral chains:

$$\text{fill}(\sum_i g_i) = \lim_{n\to\infty} \frac{\|\sum_i g_i^n\|_B}{n}$$

and then by linearity to rational chains, and by continuity to arbitrary chains in B_1. To see that a continuous extension exists, observe that for each n, there is an inequality $\|\sum_i g_i^n + \sum_j f_j^n\|_B \le \|\sum_i g_i^n\|_B + \|\sum_j f_j^n\|_B$ and therefore fill is subadditive. Since fill is homogeneous, it is evidently a class function in each argument.

With this definition, one obtains the following analogue of Lemma 2.69:

LEMMA 2.77. *For any finite linear chain* $\sum_i t_i a_i \in B_1$ *there is an equality*

$$\text{scl}(\sum_i t_i a_i) = \frac{1}{4}\text{fill}(\sum_i t_i a_i)$$

PROOF. It suffices to prove the result for integral chains; i.e. chains of the form $\sum_i a_i$ for $1 \le i \le m$.

The proof is very similar to that of Lemma 2.69; the only complication is the issue of basepoints. A surface S realizing $\text{cl}(\sum_i a_i^n)$ can be efficiently triangulated, as in Theorem 1.14, with $4\text{cl}(\sum_i a_i^n) + 3m - 4$ triangles, with exactly one vertex on each boundary component. Let T be an embedded spanning tree in the 1-skeleton, connecting up the boundary vertices (T has $m-1$ edges). We obtain a simplicial 2-complex with one vertex by collapsing T to a point, and then further collapsing degenerate triangles. Denote this 2-complex by S/T. The triangulation of S determines a triangulation of the complex S/T, with fewer triangles. Since this complex has only one vertex, it determines a (group) 2-chain A with $\|A\|_1 \le 4\text{cl}(\sum_i a_i^n) + 3m - 4$, and satisfying $\partial A = \sum_i b_i^n$ where each b_i is conjugate to a_i. Since m is fixed, and fill is a class function in each argument, as $n \to \infty$ we obtain an inequality in one direction.

Conversely, a 2-chain A with $\partial A = \sum_i a_i^n$ and $\|A\|_1$ close to $\|\sum_i a_i^n\|_B$ can be approximated by a rational chain. After multiplying through by a big integer to clear denominators one obtains an (approximating) integral chain. Gluing up triangles, one obtains a "collapsed surface" of the form S/T as above, with one vertex on each boundary component. This collapsed surface can be thickened to a genuine surface by adding a cylindrical collar to each boundary component, at the cost of adding a further $2m$ triangles. Since m is fixed but n is arbitrarily large, the desired inequality follows by applying Proposition 2.74, and Theorem 1.14. □

2.6.2. Generalized Bavard duality.

DEFINITION 2.78. Let G be a group. Let $H(G)$ (for "homogeneous") be the subspace of $B_1(G)$ spanned by elements of the form $g - hgh^{-1}$ and $g^n - ng$ for $g, h \in G$ and $n \in \mathbb{Z}$. Denote the quotient space as $B_1^H(G) := B_1(G)/H(G)$ or B_1^H for short, if G is understood.

By construction, scl vanishes on the subspace $H(G)$, and therefore descends to a pseudo-norm on B_1^H. With this notation, we obtain the following statement of generalized Bavard duality:

THEOREM 2.79 (Generalized Bavard Duality). *Let G be a group. Then for any* $\sum_i t_i a_i \in B_1^H(G)$ *there is an equality*

$$\text{scl}(\sum_i t_i a_i) = \frac{1}{2} \sup_{\phi \in Q/H^1} \frac{\sum_i t_i \phi(a_i)}{D(\phi)}$$

PROOF. The proof is the same as that of Theorem 2.70 with Lemma 2.77 in place of Lemma 2.69. □

This mixture of group theoretic and homological language is convenient for deriving some interesting corollaries.

PROPOSITION 2.80 (Finite index formula). *Let G be a group, and H a subgroup of finite index. Let $g_1, \cdots, g_m \in G$. Suppose $\pi_1(X) = G$, and let $\gamma_1, \cdots, \gamma_m$ be* *loops in X representing the conjugacy classes of the g_i. Let $p : \widehat{X} \to X$ be a finite cover corresponding to the subgroup H. Let $\beta_1, \cdots, \beta_l$ be the covers of the γ_i which lift to $\widehat{X}$, and $h_1, \cdots, h_l$ the corresponding conjugacy classes in H. Then*

$$\mathrm{scl}_G(\sum_i g_i) = \frac{1}{[G:H]} \cdot \mathrm{scl}_H(\sum_i h_i)$$

PROOF. We use Proposition 2.74. Given a map of a surface $f : (S, \partial S) \to (X, \cup_i \gamma_i)$ there is a finite covering map $\pi : (\widehat{S}, \partial\widehat{S}) \to (S, \partial S)$ such that $f\pi$ lifts to $\widehat{f} : (\widehat{S}, \partial\widehat{S}) \to (\widehat{X}, \cup_i \beta_i)$ in such a way that $p\widehat{f} = f\pi$. One way to construct such a π is to let $K < H$ be normal in G of finite index, and then take $\widehat{S}$ to be the regular cover of S corresponding to the kernel of the map $\pi_1(S) \to G/K$. Conversely, given $g : (S, \partial S) \to (\widehat{X}, \cup_i \beta_i)$ the composition pg maps S to X, wrapping the boundary around the various γ_i. The result follows. □

In the case that H is normal and g is a single element in H, the finite index formula takes the following form:

COROLLARY 2.81. *Let G be a group, and let H be a normal subgroup of finite index, with (finite) quotient group $A = G/H$. Let $h \in H$. Then*

$$\mathrm{scl}_G(h) = \frac{1}{|A|} \cdot \mathrm{scl}_H(\sum_{a \in A} aha^{-1})$$

where for each $a \in A$, the expression aha^{-1} represents the corresponding (well-defined) conjugacy class in H.

REMARK 2.82. One can give a more algebraic proof of Corollary 2.81 as follows. By Theorem 2.47, and the fact that finite groups are amenable, the map $H_b^2(G) \to H_b^2(H)$ is an isometric embedding with image equal to the A-invariant part of $H_b^2(H)$. If $\psi \in Q(H)$ then the projection ψ^A of ψ to $Q(H)^A$ is the sum $1/|A| \sum_a a^*\psi$. Here the group A acts on H by outer automorphisms: if $a = aH$ is a left coset of H, then aha^{-1} is a well-defined element of H up to an inner automorphism. In other words, $a^*\psi(h) = \phi(aha^{-1})$.

It follows that

$$\mathrm{scl}_G(h) = \sup_{\phi \in Q(G)} \frac{\phi(h)}{2D(\phi)} = \sup_{\psi \in Q(H)} \frac{\psi^A(h)}{2D(\psi^A)}$$

Now for any $\psi \in Q(H)$, one has

$$\psi^A(h) = \frac{1}{|A|} \sum_a \psi(aha^{-1}) = \frac{1}{|A|} \sum_a \psi^A(aha^{-1})$$

Furthermore, $D(\psi^A) \le D(\psi)$ by convexity. It follows that

$$\frac{1}{|A|}\mathrm{scl}_H(\sum_a aha^{-1}) = \sup_{\psi \in Q(H)} \frac{1}{|A|} \frac{\sum_a \psi(aha^{-1})}{2D(\psi)} = \sup_{\psi \in Q(H)} \frac{1}{|A|} \frac{\sum_a \psi^A(aha^{-1})}{2D(\psi^A)}$$

proving the formula.

REMARK 2.83. Corollary 2.81 is useful even (especially?) when an element $h \in H$ is in $[G, G]$ but not in $[H, H]$.

One advantage of working with the space B_1^H over B_1 is that while scl is, except in trivial cases, never a genuine norm on B_1, it is sometimes a genuine norm on B_1^H.

PROPOSITION 2.84. *Let F be a free group. Then* scl *is a genuine norm on the vector space $B_1^H(F)$.*

PROOF. A chain c in $B_1^H(F)$ has a representative of the form $\sum_i t_i w_i$ where each w_i is a cyclically reduced primitive word in F, where all coefficients t_i are nonzero, and where no two $w_i^{\pm 1}, w_j^{\pm 1}$ are conjugate for distinct i, j. After reordering, assume that the length of $w := w_1$ is at least as big as that of any w_i. Let N be a sufficiently big integer (to be determined), and let φ be the homogenization of the big Brooks counting quasimorphism H_{w^N} associated to w^N. We claim that for sufficiently big N, there is equality $\varphi(w_i) = 0$ for any $i \neq 1$. Since $\varphi(w) = 1/N$, this shows that $\text{scl}(c) \geq |t_1|/2ND(\varphi) > 0$.

To prove the claim, suppose to the contrary that for some $i \neq 1$ the infinite product w_i^∞ contains an arbitrarily big power w^N as a subword, where without loss of generality, we may assume N is positive. If $N = \text{lcm}(|w|, |w_i|)/|w|$ then w^N is conjugate to w_i^M for some M. But elements in free groups have unique primitive roots, up to conjugacy, so this implies $M = N$ and w_i is conjugate to w, contrary to hypothesis. This establishes the claim, and the proposition. □

REMARK 2.85. A similar argument using de Rham quasimorphisms in place of Brooks quasimorphisms works whenever G is equal to π_1 of a closed hyperbolic manifold. In fact, using generalized counting quasimorphisms § 3.5 one can show that scl is a norm on $B_1^H(G)$ whenever G is a hyperbolic group.

REMARK 2.86. It is not true that fill is equal to the quotient norm on B_1^H under the exact sequence

$$H \to B_1 \to B_1^H$$

where B_1 and H have the $\|\cdot\|_B$ norm. For instance, in a free group, a (nontrivial) commutator $ghg^{-1}h^{-1}$ has scl norm 1/2, and therefore fill norm 2. On the other hand, the chains $ghg^{-1}h^{-1}$ and $ghg^{-1}h^{-1} + hgh^{-1} - g$ differ by an element of H, and

$$\partial(ghg^{-1}h^{-1}, hgh^{-1}) = ghg^{-1}h^{-1} + hgh^{-1} - g$$

so $\|ghg^{-1}h^{-1} + hgh^{-1} - g\|_B \leq 1$.

2.7. Further properties

In this section we enumerate some further properties of scl which will be used in the sequel.

2.7.1. Extremal quasimorphisms. Theorem 2.70 provides a method of calculating scl in some cases, especially when the dimension of $Q(G)$ is small. Given an element $a \in [G, G]$, it is natural to ask whether the supremum of $\phi(a)/D(\phi)$ is realized by some $\phi \in Q(G)$.

DEFINITION 2.87. Let $a \in [G, G]$. An element $\phi \in Q(G)$ is *extremal* for a if

$$\text{scl}(a) = \frac{\phi(a)}{2D(\phi)}$$

The union of 0 with the set of homogeneous quasimorphisms on G which are extremal for a is denoted $Q_a(G)$.

The next Proposition shows that extremal quasimorphisms always exist.

PROPOSITION 2.88. *Let $a \in [G, G]$. Then $Q_a(G)$ is a nontrivial convex cone in $Q(G)$ which is closed both in the defect and the weak* topology.*

Proof. Recall from Remark 2.60 that there is an isomorphism of vector spaces $Q/H^1 \cong (Z_2^{l_1}/\overline{Z}_2)'$. As a dual space, we can endow Q/H^1 with the weak* topology. A subset closed in the weak* topology is also closed in the defect topology.

The space

$$K := \{\varphi \in Q/H^1 \text{ such that } D(\phi) \le 1/2\}$$

is convex, closed and bounded with respect to the defect norm and therefore also with respect to the operator norm (since these two norms differ by a factor of at most 2). Hence K is weak* compact.

Fix an element $a \in [G,G]$ and for each n, define

$$K^n := \{\varphi \in K \text{ such that } \varphi(a) \ge \text{scl}(a) - 1/n\}$$

Let us show that K^n is weak* closed. Since $[G,G] \subset B_1$, there is $A \in C^2$ such that $dA = a$. The element $A - \sigma(a)$, where σ is the section defined in Lemma 2.61, satisfies $A - \sigma(a) \in Z_2^{l_1}$ and further satisfies $\delta\varphi(A - \sigma(a)) = \varphi(a)$ for any homogeneous φ. This together with the defining property of K^n shows that K^n is weak* closed.

The K^n are closed and contained in K and are therefore weak* compact. By Bavard duality (Theorem 2.70), each K^n is nonempty, and therefore their intersection is nonempty. Any element $\varphi \in \cap_n K^n$ has $\varphi(a) = \text{scl}(a)$ and $D(\varphi) = 1/2$. Conversely any $\varphi \in Q_a(G)$ can be scaled to have $D(\varphi) = 1/2$, and therefore $Q_a(G)$ is exactly equal to the cone on the weak* compact set $\cap_n K^n$. This completes the proof. □

Remark 2.89. In a similar way we may define $Q_a(c)$ for any chain $c \in B_1$. The proof of Proposition 2.88 extends easily to this case.

2.7.2. Left exactness and Bouarich's Theorem. For the convenience of the reader, we provide a proof of Bouarich's Theorem 2.49. Recall that Bouarich's Theorem says if

$$K \xrightarrow{\iota} G \xrightarrow{\rho} H \to 0$$

is an exact sequence of groups then the induced sequence

$$0 \to H_b^2(H;\mathbb{R}) \xrightarrow{\rho^*} H_b^2(G;\mathbb{R}) \xrightarrow{\iota^*} H_b^2(K;\mathbb{R})$$

is left exact. In fact, it is no more difficult to give a proof of Bouarich's theorem which is valid for any Abelian coefficient group; in particular, the proof we give below applies to bounded cohomology with $\mathbb{Z}$ coefficients.

Proof. Without loss of generality, we can replace K by its image $\iota(K)$. So we can assume K is a subgroup of G, and ι is the inclusion homomorphism. Since $\rho\iota$ is the zero map, the composition $H_b^2(H) \to H_b^2(G) \to H_b^2(K)$ is zero. So we just need to check that ρ^* is injective, and that everything in $\ker(\iota^*)$ is in the image of the map ρ^*.

Claim. *The map $\rho^* : H_b^2(H) \to H_b^2(G)$ is an injection.*

Proof. Suppose ψ be a bounded 2-cocycle on H whose image in $H_b^2(H)$ is nonzero, but for which $\rho^*\psi = \delta\phi$ on G, where ϕ is bounded. Observe that for all $a_1, a_2 \in G$ and $k_1, k_2 \in K$ that

$$\phi(a_1) + \phi(a_2) - \phi(a_1 a_2) = \phi(a_1 k_1) + \phi(a_2 k_2) - \phi(a_1 k_1 a_2 k_2)$$

In particular, $\phi(k^{n+1}) - \phi(k^n) = \phi(ak^{n+1}) - \phi(ak^n)$ for any $a \in G$, $k \in K$. Taking $a = k$ this implies $\phi(k^n) = n(\phi(k) - \phi(\text{id})) + \phi(\text{id})$. But ϕ is bounded, so $\phi(k) -$

$\phi(\text{id}) = 0$ for all $k \in K$, and more generally, ϕ is constant on left cosets. This implies that ϕ descends to a bounded function ϕ_H on $H = G/K$ which by construction satisfies $\delta\phi_H = \psi$. □

CLAIM. *Let $[\psi] \in H_b^2(G)$ be in the kernel of $\iota^* : H_b^2(G) \to H_b^2(K)$. Then $[\psi]$ is in the image of $H_b^2(H)$.*

PROOF. By hypothesis, for any representative ψ of $[\psi]$ there is a bounded function ϕ on K such that $\delta\phi = \psi$ on K. If $\psi(\text{id}, \text{id}) = c \neq 0$ then we replace ψ by $\psi - \delta h_c$ where h_c is the constant bounded 1-cochain $h_c(g) = c$. So without loss of generality, we can assume that $\psi(\text{id}, \text{id}) = 0$. This leads to the convenient normalization $\phi(\text{id}) = 0$.

We want to extend ϕ in a suitable way to a function ϕ_G on all of G. For each $h_i \in H$, choose a left coset representative g_i of h_i in G. For each h_i we define $\phi_G(g_i) = 0$. Then for each $k \in K$ we set $\phi_G(g_i k) = \psi(g_i, k) - \phi(k)$. Since ϕ and ψ are bounded, ϕ_G is bounded. Now define $\psi' = \psi + \delta\phi_G$. Since ϕ_G is bounded, ψ' and ψ represent the same cohomology class. Moreover, for any g in G and $k \in K$ we write $g = g_i k_i$ and calculate

$$\begin{aligned}\psi'(g, k) &= \psi(g_i k_i, k) + \phi_G(g_i k_i) + \phi_G(k) - \phi_G(g_i k_i k) \\ &= \psi(g_i k_i, k) + \psi(g_i, k_i) - \phi(k_i) + \psi(\text{id}, k) - \phi(k) - \psi(g_i, k_i k) + \phi(k_i k)\end{aligned}$$

Since $\phi(\text{id}) = 0$, we have $\psi(\text{id}, k) = \delta\phi(\text{id}, k) = \phi(\text{id}) + \phi(k) - \phi(k) = 0$. Moreover, $-\phi(k_i) - \phi(k) + \phi(k_i k) = -\delta\phi(k_i, k) = -\psi(k_i, k)$. Therefore we can write

$$\psi'(g, k) = \psi(g_i k_i, k) + \psi(g_i, k_i) - \psi(k_i, k) - \psi(g_i, k_i k) = -\delta\psi(g_i, k_i, k) = 0$$

We claim that ψ' can be obtained by pulling back a bounded 2-cocycle from H. Let $g_1, g_2 \in G$ and $k \in K$. Since $\delta\psi'(g_1, g_2, k) = 0$, we calculate

$$\psi'(g_1, g_2 k) - \psi'(g_1, g_2) = \psi'(g_1 g_2, k) - \psi'(g_2, k) = 0$$

and therefore $\psi'(g_1, g_2 k) = \psi'(g_1, g_2)$ for any $g_1, g_2 \in G$ and any $k \in K$.

Similarly, since $\delta\psi'(g_1, k, g_2) = 0$ we have

$$\psi'(g_1, k g_2) - \psi'(g_1, k) = \psi'(g_1 k, g_2) - \psi'(k, g_2)$$

We have shown that $\psi'(g_1, k) = 0$. Moreover, $\psi'(g_1, k g_2) = \psi'(g_1, g_2(g_2^{-1} k g_2))$ which is equal to $\psi'(g_1, g_2)$ by our earlier calculation. Rearranging, we obtain

$$\psi'(g_1 k, g_2) - \psi'(g_1, g_2) = \psi'(k, g_2)$$

and therefore

$$\psi'(g_1 k^n, g_2) = \psi'(g_1, g_2) + n\psi'(k, g_2)$$

for any integer n. Since n is arbitrary but ψ' is bounded, we see that $\psi'(k, g_2) = 0$ for any $g_2 \in G$ and $k \in K$ and therefore also $\psi'(g_1 k, g_2) = \psi'(g_1, g_2)$. In particular, ψ' is constant on left cosets of K, and descends to a cocycle on H. □

This completes the proof of Bouarich's Theorem. □

REMARK 2.90. A similar but more straightforward argument proves the left exactness of Q.

REMARK 2.91. There is a more direct proof of Bouarich's Theorem using spectral sequences. In fact, the astute reader will recognize that the proof given above is really a spectral sequences argument in disguise, together with the observation that H_b^1 is always zero. However one must be careful in general, since bounded cohomology is typically not separated in degree 3 and higher (see the end of § 2.4.1 and § 2.5.1). This is a point which is sometimes overlooked in the literature on bounded cohomology. Nevertheless, in sufficiently low dimensions, such an argument can be made to work. See e.g. Chapter 12 of [**157**], especially Example 12.4.3.

2.7.3. Rotation numbers. As an application of Theorem 2.70 we obtain a precise estimate of the defect of rotation number.

PROPOSITION 2.92. *Let G be a subgroup of* $\mathrm{Homeo}^+(S^1)$ *and let $\widehat{G}$ be the preimage in* $\mathrm{Homeo}^+(\mathbb{R})$. *Then $D(\mathrm{rot}) \le 1$ as a homogeneous quasimorphism on $\widehat{G}$.*

PROOF. For the sake of brevity, let $T = \mathrm{Homeo}^+(S^1)$ and let $\widehat{T} = \mathrm{Homeo}^+(\mathbb{R})^{\mathbb{Z}}$. By Remark 2.44 we see that $Q(T) = 0$. The exact sequence $\mathbb{Z} \to \widehat{T} \to T$ together with Bouarich's Theorem 2.49 and the vanishing of H_b^* for amenable groups implies that $H_b^2(T) \to H_b^2(\widehat{T})$ is an isomorphism. On the other hand, the map $H^2(T) \to H^2(\widehat{T})$ is not injective, and the kernel is generated by the class of the (universal) central extension $\widehat{T} \to T$. It follows that $Q(\widehat{T})$ is 1-dimensional, and generated exactly by rot. By Theorem 2.43 there is an equality $\mathrm{scl}(a) = |\mathrm{rot}(a)|/2$ for every $a \in \widehat{T}$ and therefore $D(\mathrm{rot}) = 1$, by Bavard's Theorem 2.70. It follows that $D(\mathrm{rot}) \le 1$ on any subgroup of $\widehat{T}$. □

2.7.4. Free products. Bavard Prop. 3.7.2 [**8**] asserts that if G_1 and G_2 are two groups, and $G = G_1 * G_2$ is their free product, then for all nontrivial elements $g_i \in G_i$, there is an equality $\mathrm{scl}(g_1 g_2) = \mathrm{scl}(g_1) + \mathrm{scl}(g_2) + 1/2$. This assertion is not quite true as stated. Nevertheless, it turns out that Bavard's assertion is true *when g_1 and g_2 have infinite order*, and can be suitably modified when one or both of them are torsion. We give the correct statement and proof, and defer a discussion of Bavard's argument and what can be salvaged from it to the sequel.

THEOREM 2.93 (Product formula). *Let G_1, G_2 be groups, and for $i = 1, 2$ let g_i be a nontrivial element in G_i of order n_i. Let $G = G_1 * G_2$. Then there is an equality*

$$\mathrm{scl}_G(g_1 g_2) = \mathrm{scl}_{G_1}(g_1) + \mathrm{scl}_{G_2}(g_2) + \frac{1}{2}\left(1 - \frac{1}{n_1} - \frac{1}{n_2}\right)$$

where $1/n_i$ may be replaced by 0 when $n_i = \infty$.

PROOF. Build a space X as follows. Let X_1, X_2 be spaces with $\pi_1(X_i) = G_i$, and let γ_i be a loop in X_i representing the conjugacy class of g_i. Let P be a pair of pants. Let $X = X_1 \cup X_2 \cup P$ be obtained by gluing two boundary components of P to γ_1 and γ_2 respectively, and let γ_P denote the unglued boundary component of P.

Let S be a surface with one boundary component, and $f : S \to X$ a map sending ∂S to γ_P with degree n. We have $\mathrm{scl}(g_1 g_2) \le -\chi(S)/2n$. Make f transverse to the γ_i. The surface is decomposed into *pieces*, which are the closures, in the path topology, of $S - f^{-1}(\gamma_1 \cup \gamma_2)$. We say that f is *efficient* if no piece has a boundary component which maps with degree zero to a γ_i, and if no piece is an annulus with both boundary components mapping to the same γ_i with opposite degree.

If S is not efficient, the Euler characteristic of S can be increased by surgering S along a circle which maps to some γ_i with degree 0 (and is therefore null-homotopic), or simplified by homotoping a trivial annulus. So without loss of generality, it suffices to consider the case that f is efficient. Let S_i denote the union of the pieces mapping to X_i, and S_P the union of pieces mapping to P. Let f_1, f_2, f_P be the restrictions of f to these unions. These maps are all proper. Since f is efficient, no piece mapping to P is a disk or annulus. In other words, S_P admits a hyperbolic metric. Moreover, the only disk pieces are components of S_i mapping with degree a multiple of n_i to γ_i, in the case g_i is torsion.

Since f_P is proper, it has a well-defined degree. Since $f_P^{-1}(\gamma_P)$ is equal to ∂S, the degree is n. By the definition of degree, the union of components of ∂S_P mapping to each γ_i maps with degree n, and therefore $n(S_1) = n(S_2) = n$ in the notation of Proposition 2.10. By replacing f_P by a pleated map (with respect to a choice of hyperbolic structures on S_P and on P) and Gauss–Bonnet, we obtain an inequality $-\chi(S_P)/2n \ge -\chi(P)/2 = 1/2$.

If each g_1, g_2 has infinite order, no component of S_i is a disk. In this case, $-\chi^-(S) = -\chi^-(S_1) - \chi^-(S_2) - \chi^-(S_P)$, and therefore

$$\frac{-\chi^-(S)}{2n} = \frac{-\chi^-(S_1)}{2n} + \frac{-\chi^-(S_2)}{2n} + \frac{-\chi^-(S_P)}{2n} \ge \mathrm{scl}(g_1) + \mathrm{scl}(g_2) + \frac{1}{2}$$

Since S was arbitrary, we obtain an inequality

$$\mathrm{scl}(g_1 g_2) \ge \mathrm{scl}(g_1) + \mathrm{scl}(g_2) + \frac{1}{2}$$

Conversely, by the proof of Lemma 2.24 the elements $(g_1 g_2)^{2n}$ and $g_1^{2n} g_2^{2n}$ differ by at most n commutators, and therefore we obtain the first inequality

$$\mathrm{scl}(g_1 g_2) \le \mathrm{scl}(g_1) + \mathrm{scl}(g_2) + \frac{1}{2}$$

This proves the theorem when the g_i have infinite order.

If g_i is torsion of order n_i, then S_i may have disk components whose boundaries map to g_i with degree a multiple of n_i. In this case, S_i might have as many as n/n_i disk components, and therefore $\chi(S_i)$ might be as big as n/n_i, so we obtain an inequality

$$-\chi^-(S) \ge -\chi^-(S_1) - \chi^-(S_2) - \chi^-(S_P) - \frac{n}{n_1} - \frac{n}{n_2}$$

which, after dividing by $2n$, and taking the infimum over all S, gives

$$\mathrm{scl}(g_1 g_2) \ge \mathrm{scl}(g_1) + \mathrm{scl}(g_2) + \frac{1}{2}\left(1 - \frac{1}{n_1} - \frac{1}{n_2}\right)$$

To obtain the reverse inequality, replace P by an orbifold with a cone point of order n_i in place of the γ_i boundary component(s) and take a finite cover which is a smooth surface. This completes the proof. □

REMARK 2.94. The use of geometric language is really for convenience of exposition rather than mathematical necessity. A similar argument could be made by replacing maps to X with equivariant maps to a suitable Bass–Serre tree.

One drawback of the method of proof is that it does not exhibit an extremal homogeneous quasimorphism for the element $g_1 g_2$. In the next section we show how to construct such an extremal quasimorphism in the case that G_1 and G_2 are *left orderable*.

REMARK 2.95. Bavard, in [**8**], exhibits a nontrivial quasimorphism for g_1g_2 arising from the structure of $G_1 * G_2$ as a free product and its action on a Bass–Serre tree, which is a special case of a construction that will be discussed in more detail in § 3.5. One can estimate the defect of the quasimorphism constructed in this way, but the estimate is not good enough to establish Theorem 2.93.

2.7.5. Left-orderability.

DEFINITION 2.96. Let G be a group. G is *left orderable* (LO for short) if there is a total ordering $<$ on G which is invariant under left multiplication. That is, for all $a, b, c \in G$, the inequality $a < b$ holds if and only if $ca < cb$.

Right orderability is defined similarly. A group is left orderable if and only if it is right orderable. The difference is essentially psychological.

EXAMPLE 2.97 (Locally indicable). A group is *locally indicable* if every nontrivial finitely generated subgroup admits a surjective homomorphism to $\mathbb{Z}$. For example, free groups are locally indicable. A more nontrivial example, due to Boyer–Rolfsen–Wiest [**22**] says that if M is an irreducible 3-manifold, and $H^1(M) \neq 0$ then $\pi_1(M)$ is locally indicable.

A theorem of Burns–Hale [**36**] says that every locally indicable group is left orderable.

Left orderability is intimately bound up with 1-dimensional dynamics. The following "folklore" theorem is very well-known.

THEOREM 2.98 (Action on $\mathbb{R}$). *A countable group G is left orderable if and only if there is an injective homomorphism $G \to \text{Homeo}^+(\mathbb{R})$.*

We give a sketch of a proof. For more details, see [**40**].

PROOF. Suppose G acts faithfully on $\mathbb{R}$ by homeomorphisms. Suppose $p \in \mathbb{R}$ has trivial stabilizer. Then define $a > \text{id}$ if and only if $a(p) > p$. Conversely, suppose G is left orderable. The order topology on G makes G order-isomorphic to a countable subset of $\mathbb{R}$. Include $G \hookrightarrow \mathbb{R}$ in an order-preserving way, compatibly with the order topology. Then the action of G on itself extends to an action on the closure of its image. The complement is a countable union of intervals; the action of G extends uniquely to a permutation action on these intervals. □

The first part of the next proposition is a special case of Theorem 2.93; however, the proof is different, and shows how to construct an explicit extremal quasimorphism for g_1g_2.

PROPOSITION 2.99. *Let G_1, G_2 be left orderable, and suppose $g_i \in G_i$ are nontrivial. Then there is an equality*

$$\text{scl}(g_1g_2) = \text{scl}(g_1) + \text{scl}(g_2) + \frac{1}{2}$$

Moreover there is an explicit construction of an extremal quasimorphism for g_1g_2 in terms of extremal quasimorphisms for g_1 and g_2.

PROOF. Assume first that G_1, G_2 are countable. Using Theorem 2.98, construct an orientation-preserving action of $G_1 * G_2$ on S^1 where G_1 fixes the point -1 and G_2 fixes the point 1 (here we think of S^1 as the unit circle in $\mathbb{C}$). Since g_1, g_2 are nontrivial, without loss of generality we can assume $g_1(i) = -i$ and $g_2(-i) = i$.

But then g_1g_2 has a fixed point, and therefore its rotation number is trivial (in $\mathbb{R}/\mathbb{Z}$). We lift the action to an action on $\mathbb{R}$, which can be done by lifting each G_i individually to have a global fixed point. Then rot is a homogeneous quasimorphism on $G_1 * G_2$, which vanishes on G_1 and on G_2, and satisfies $\text{rot}(g_1g_2) = 1$. Proposition 2.92 shows that $D(\text{rot}) \le 1$. Adding to rot pullbacks of extremal quasimorphisms with defect 1 for g_1 and g_2 under the surjections $G_1 * G_2 \to G_1$ and $G_1 * G_2 \to G_2$, one obtains an explicit extremal quasimorphism for g_1g_2 which, by Bavard duality, proves the proposition.

If G_1, G_2 are not countable, one substitutes actions on circularly ordered sets for actions on circles. The distinction between these two contexts is more psychological than substantial. See e.g. [**40**], especially Chapter 2, for a discussion. □

Example 2.100 (Bavard, p. 146 [**8**]). In $F_2 = \langle u, v\rangle$ the element $[u, v]$ satisfies $\text{scl}([u, v]) = 1/2$, by Theorem 1.14 and Theorem 2.70. Let $G = \langle u_1, v_1, \cdots, u_k, v_k\rangle$. Then by Proposition 2.99 and induction,

$$\text{scl}(\prod_i [u_i, v_i]^{p_i}) = \frac{1}{2}\sum |p_i| + \frac{k-1}{2}$$

since free groups are locally indicable and therefore left orderable (see Example 2.97).

The interaction of left orderability and scl (especially in order to obtain sharp estimates in free groups) will be discussed again in § 4.3.4.

2.7.6. Self-products. There is an analogue of Theorem 2.93 with (free) HNN extensions in place of free products. For convenience, we state and prove the theorem only in the case that the elements in question are torsion free.

Theorem 2.101 (Self-product formula). *Let G be a group, and $g_1, g_2 \in G$ two elements of infinite order. Let $G' = G * \langle t\rangle$. Then there is an equality*

$$\text{scl}_{G'}(g_1tg_2t^{-1}) = \text{scl}_G(g_1 + g_2) + \frac{1}{2}$$

Proof. Let X be a space with $\pi_1(X) = G$. Let γ_1, γ_2 be loops representing the conjugacy classes of g_1, g_2 respectively. Let P be a pair of pants, and let $Y = X \cup P$ be obtained by gluing two boundary components of P to γ_1 and γ_2 respectively, and let γ_P denote the unglued boundary component of P.

Notice that $\pi_1(Y) = G'$ and γ_P represents the conjugacy class of $g_1tg_2t^{-1}$. If $f : S \to Y$ sends ∂S to γ_P with degree n, then after making f efficient, S decomposes into $f_X : S_X \to X$ and $f_P : S_P \to P$. The degree of f_P is n, so $-\chi^-(S_P)/2n \ge 1/2$, and $-\chi^-(S_X)/2n$ is an upper bound for $\text{scl}(g_1 + g_2)$. Since the g_i have infinite order, no component of S_X is a disk, and therefore $-\chi^-(S) = -\chi^-(S_P) - \chi^-(S_X)$. The proof now follows, as in the proof of Theorem 2.93, from Proposition 2.10 and Proposition 2.74. □

Remark 2.102. Note that the same proof shows

$$\text{scl}_{G'}(g_1tg_2t^{-1} + \sum t_ig_i) = \text{scl}_G(g_1 + g_2 + \sum t_ig_i) + \frac{1}{2}$$

for any $\sum t_ig_i \in B_1^H$ where we sum over $i \ge 3$.

Remark 2.103. By Remark 2.102 and by the linearity and continuity of scl on B_1^H, the calculation of scl on B_1^H can be reduced to calculations of scl on "ordinary" elements of $G * F$ for sufficiently large free groups F.

2.7.7. LERF and injectivity. Recall Proposition 2.10, which says that if X is a space with $\pi_1(X) = G$, and γ is a loop in X representing the conjugacy class of a, then

$$\mathrm{scl}(a) = \inf_S \frac{-\chi^-(S)}{2n(S)}$$

where the infimum is taken over all maps of oriented surfaces $f : S \to X$ whose boundary components all map to γ with sum of degrees equal to $n(S)$. Recall (Definition 2.11) that f, S is said to be *extremal* if it realizes the infimum. The following proposition says that extremal surfaces must be π_1-injective.

Proposition 2.104 (injectivity). *Let X, γ be as above. Suppose f, S as above is extremal. Then the map $f : S \to X$ induces a monomorphism $\pi_1(S) \to \pi_1(X)$.*

Before we prove the proposition, we must discuss the property LERF for surface groups.

Definition 2.105. Let G be a group. Then G is *locally extended residually finite* (or LERF for short) if all of its finitely generated subgroups are separable. That is, for all finitely generated subgroups H and all elements $a \in G - H$ there is a subgroup H' of G of finite index which contains H but not a.

Example 2.106 (Malcev; polycyclic groups). A solvable group is *polycyclic* if all its subgroups are finitely generated. Malcev [**142**] showed that polycyclic groups are LERF.

Example 2.107 (Hall; free groups). Marshall Hall [**102**] showed that free groups are LERF. In fact, he showed that free groups satisfy the stronger property that finitely generated subgroups are *virtual retracts*. We sketch an illuminating topological proof of this fact due to Stallings [**191**].

Let F be free, and let G be a finitely generated proper subgroup. Represent $F = \pi_1(X)$ where X is a wedge of circles, and let $\tilde{X}$ be a cover of X corresponding to the subgroup G. Since G is a finitely generated subgroup of a free group, it is free of finite rank, so $\tilde{X}$ deformation retracts to a compact subgraph X_G with $\pi_1(X_G) = G$. Each directed edge of X_G is labeled by a generator of F. Let X'_G be another copy of X_G with each directed edge labeled by the inverse of the corresponding label in X_G. For each vertex v of X_G, let v' be the corresponding vertex of X'_G. Join v to v' by a collection of edges, one for each generator of $\pi_1(X)$ not represented by an edge in X_G with a vertex at v. Let the result be X''_G. Then by construction, X''_G is a finite covering of X, and therefore corresponds to a finite index subgroup H of F. Moreover, by construction, G is a free summand of H.

Example 2.108 (Scott; surface groups). Peter Scott [**185**] showed that surface groups are LERF. For surfaces with boundary, this is a special case of Example 2.107, but even in this case, Scott's proof is different and illuminating.

Let S be a surface with $\chi(S) < 0$. Observe that S can be tiled by right-angled hyperbolic pentagons, for some choice of hyperbolic structure on S. Let G be a finitely generated subgroup of $\pi_1(S)$, and let $\tilde{S}$ be the covering corresponding to G. The surface $\tilde{S}$ deformation retracts to a compact subsurface X with $\pi_1(X) = G$. This subsurface can be engulfed by a convex union Y of right-angled hyperbolic pentagons. Since all the pentagons are right-angled, Y is a surface with right-angled corners. There is a hyperbolic orbifold obtained from Y by adding mirrors to the non-boundary edges. This orbifold has a finite index subgroup, containing G, which is also finite index in $\pi_1(S)$.

A geometric corollary of property LERF for free and surface groups is the fact that for any hyperbolic surface S and any geodesic loop γ in S there is a finite cover $\tilde{S}$ of S to which γ lifts as an embedded loop. Using this fact, we now prove Proposition 2.104:

PROOF. Suppose S minimizes $-\chi(S)/2n(S)$ but $f : S \to X$ is not injective in π_1. Let $a \in \pi_1(S)$ be in the kernel. Choose a hyperbolic structure on S, and represent the conjugacy class of a by a geodesic loop γ in S. If γ is embedded, compress S along γ to produce a surface S' satisfying $-\chi(S') < -\chi(S)$. The compression factors through f, and there is a map $f' : S' \to X$ satisfying $n(S') = n(S)$, contrary to the minimality of S.

If γ is not embedded, let $\tilde{S}$ be a finite cover of S to which γ lifts as an embedded loop. Let $\pi : \tilde{S} \to S$ be the covering map. Since both χ and $n(\cdot)$ are multiplicative under covers, there is an equality $-\chi(S)/2n(S) = -\chi(\tilde{S})/2n(\tilde{S})$. But $\tilde{S}$ can be compressed along γ to produce a new surface $\tilde{S}'$. The compression factors through $f\pi$, contradicting the minimality of S, as before. This contradiction shows that f is injective on $\pi_1(S)$, as claimed. □

This lets us give a short proof of the following corollary. Note that this corollary is easy to prove in many other ways. For instance, it follows from the fact that every subgroup of a free group is free, and from the theorem of Malcev [**141**] that free groups are Hopfian (i.e. surjective self-maps are injective).

COROLLARY 2.109. *Let $\rho : F_2 \to F$ be a homomorphism from F_2, the free group on two elements, to F, a free group. If the image is not Abelian, ρ is injective.*

PROOF. Let X be a wedge of circles with $\pi_1(X) = F$. Let $F_2 = \langle a, b\rangle$. The map ρ defines a map from a punctured torus S into X, taking the boundary to $\rho([a,b])$. By hypothesis, this element is nontrivial in F. If ρ is not injective, Proposition 2.104 implies $\text{scl}(\rho([a,b])) < 1/2$. But we will show in § 4.3.4 that every nontrivial element in a free group satisfies $\text{scl} \geq 1/2$. □

CHAPTER 3

Hyperbolicity and spectral gaps

There are two main sources of quasimorphisms: *hyperbolic geometry* (i.e. negative curvature) and *symplectic geometry* (i.e. partial orders and causal structures). In this chapter we study scl in hyperbolic manifolds, and more generally, in word-hyperbolic groups in the sense of Rips and Gromov [**98**] and groups acting on hyperbolic spaces (we return to symplectic geometry, and quasimorphisms with a dynamical or causal origin in Chapter 5). The construction of explicit quasimorphisms is systematized by Bestvina–Fujiwara ([**13**]), who show that in order to construct (many) quasimorphisms on a group G, it suffices to exhibit an isometric action of G on a δ-hyperbolic space X which is *weakly properly discontinuous* (see Definition 3.51). It is crucial for many important applications that X need not be itself proper.

The relationship between negative curvature and quasimorphisms is already evident in the examples from § 2.3.1. If M is a closed hyperbolic manifold, the space of smooth 1-forms $\Omega^1 M$ injects into $Q(\pi_1(M))$. Evidently, quasimorphisms are sensitive to a great deal of the geometry of M; one of the goals of this chapter is to sharpen this statement, and to say what kind of geometry quasimorphisms are sensitive to.

A fundamental feature of the geometry of hyperbolic manifolds is the *thick-thin decomposition*. In each dimension n there is a universal constant $\epsilon(n)$ (the *Margulis constant*) such that the part of a hyperbolic n-manifold M with injectivity radius less than ϵ (i.e. the "thin" piece) has very simple topology — each component is either a neighborhood of a cusp, or a tubular neighborhood of a single short embedded geodesic. Margulis' observation implies that in each dimension, there is a *universal* notion of what it means for a closed geodesic to be *short*.

In this chapter we prove fundamental inequalities relating length to scl in hyperbolic spaces and to show that there is a *universal* notion of what it means for a conjugacy class in a hyperbolic group to have small scl. We think of this as a kind of *homological Margulis Lemma*. These inequalities generalize to certain groups acting on hyperbolic spaces, such as amalgamated free products and mapping class groups of surfaces.

Much of the content in this chapter is drawn from papers of Bestvina, Calegari, Feighn, and Fujiwara (sometimes in combination), especially [**82, 83, 13, 42, 49, 12**].

3.1. Hyperbolic manifolds

We start with the simplest and most explicit examples of groups acting on hyperbolic spaces, namely fundamental groups of hyperbolic manifolds. In this context, scl can be controlled by directly studying maps of surfaces to manifolds.

When we come to study more general hyperbolic spaces, the use of quasimorphisms becomes more practical.

3.1.1. Margulis' Lemma. The most straightforward formulation of Margulis' Lemma is the following:

THEOREM 3.1 (Margulis' Lemma [**123**]). *For each dimension n there is a positive constant $\epsilon(n)$ (called a* Margulis constant*) with the following property. Let Γ be a discrete subgroup of* $\text{Isom}(\mathbb{H}^n)$. *For any $x \in \mathbb{H}^n$ the subgroup $\Gamma_x(\epsilon)$ of Γ generated by elements which translate x less than ϵ is virtually Abelian.*

Here a group is said to *virtually* satisfy some property P if it contains a subgroup of finite index which satisfies P. If Γ is torsion-free, $\Gamma_x(\epsilon)$ is free Abelian. If Γ is cocompact, $\Gamma_x(\epsilon)$ is either trivial or isomorphic to $\mathbb{Z}$. If M is a hyperbolic manifold, there is a so-called *thick-thin* decomposition of M into the *thin part*, namely the subset $M_{<\epsilon}$ consisting of points where the injectivity radius is less than ϵ, and the *thick part*, namely the subset $M_{\geq\epsilon}$ which is the complement of $M_{<\epsilon}$. Margulis' Lemma implies that if M is complete with finite volume, $M_{<\epsilon}$ is a disjoint union of cusps and solid torus neighborhoods of short simple geodesics.

REMARK 3.2. Good estimates for $\epsilon(n)$ are notoriously difficult to obtain. In dimension 2 there is an elementary estimate $\epsilon(2) \geq \text{arcsinh}(1) = 0.8813\cdots$ due to Buser [**37**]. Meyerhoff [**152**] showed $\epsilon(3) \geq 0.104$, and Kellerhals [**124, 125**] showed $\epsilon(n) \geq \sqrt{3}/9\pi = 0.0612\cdots$ for $n = 4, 5$, and obtained an explicit estimate [**126**] for arbitrary n:

$$\epsilon(n) \geq \frac{2}{3^{\nu+1}\pi^{\nu}} \int_0^{\pi/2} \sin^{\nu+1} t dt$$

where $\nu = [\frac{n-1}{2}]$. The same paper gives explicit lower bounds on the diameter of an embedded tube around a sufficiently short geodesic.

3.1.2. Drilling and Filling. In 3-dimensions, Margulis' Lemma implies that a sufficiently short geodesic is simple, and it can be *drilled* to produce a cusped hyperbolic 3-manifold. That is, the open manifold $M - \gamma$ admits a complete finite-volume hyperbolic structure, defining a hyperbolic manifold M_γ. We denote this suggestively by

$$M \xrightarrow{\text{drill}} M_\gamma$$

Conversely, M can be obtained from M_γ by adding a solid torus under hyperbolic Dehn surgery

$$M_\gamma \xrightarrow{\text{fill}} M$$

The geodesic γ is the core of the added solid torus.

Let $T = \partial N(\gamma)$ be the torus cusp of M_γ. Choose meridian-longitude generators m, l for $H_1(T; \mathbb{Z})$ so that the longitude is trivial in $H_1(M_\gamma; \mathbb{Q})$, and the meridian intersects the longitude once. Note that the meridian is ambiguous, and different choices differ by multiples of the longitude.

Some multiple n of the longitude l is trivial in $H_1(M_\gamma; \mathbb{Z})$ and bounds a surface S. Let p and q be coprime integers, and let $M_{p/q}$ be the result of p/q Dehn surgery on M_γ; i.e. topologically, $M_{p/q}$ is obtained from M_γ by adding a solid torus in such a way that the meridian of the added solid torus represents a primitive class $pm+ql$ in $H_1(T; \mathbb{Z})$. The p, q co-ordinates depend on the choice of meridian m. A change of basis $m \to m + l$ induces $p \to p$ and $q \to q - p$.

Thurston's hyperbolic Dehn surgery Theorem ([**198, 10**]) says that except for finitely many choices of p/q, the manifold $M_{p/q}$ is hyperbolic. Moreover, as p or q or both go to infinity, $\text{length}(\gamma) \to 0$, and the geometry of $M_{p/q}$ converges on compact subsets (in the Gromov–Hausdorff sense) to that of M_γ.

The longitude wraps p times around the core γ of the added solid torus in $M_{p/q}$. Hence ∂S wraps np times. If a denotes the conjugacy class in $\pi_1(M_{p/q})$ corresponding to the free homotopy class of γ, we obtain an estimate

$$\text{scl}(a) \le \frac{-\chi(S)}{2np}$$

In particular, for fixed M_γ, and for any positive constant δ, away from finitely many lines in Dehn surgery space (corresponding to choices p, q for which $|p|$ is small) the core of the added solid torus has scl $< \delta$. Heuristically, *most sufficiently short geodesics in hyperbolic* 3*-manifolds have arbitrarily small* scl.

Conversely, we will see that conjugacy classes in $\pi_1(M)$ with *sufficiently small* scl are represented by arbitrarily short geodesics.

3.1.3. Pleated surfaces. To study scl in $\pi_1(M)$, we need to probe M topologically by maps of surfaces $S \to M$. Under suitable geometric hypotheses, it makes sense to take representative maps of surfaces which are tailored to the geometry of M. For M hyperbolic, a very useful class of maps of surfaces into M are so-called *pleated surfaces*.

Pleated surfaces were introduced by Thurston [**198**].

DEFINITION 3.3. Let M be a hyperbolic manifold. A *pleated surface* is a complete hyperbolic surface S of finite area, together with an isometric map $f : S \to M$ which takes cusps to cusps, and such that every $p \in S$ is in the interior of a straight line segment which is mapped by f to a straight line segment.

Note that the term "isometric map" here means that f takes rectifiable curves on S to rectifiable curves in M of the same length.

The set of points $L \subset S$ where the line segment through p is unique is called the *pleating locus*. It turns out that L is a *geodesic lamination* on S; i.e. a closed union of disjoint simple geodesics. Moreover, the restriction of f to each component of $S - L$ is totally geodesic.

Since S has finite area, L is nowhere dense, and $S - L$ has full measure in S.

EXAMPLE 3.4 (Thurston's spinning construction; § 8.8 and § 8.10 [**198**]). The most important and useful method of producing pleated surfaces is Thurston's *spinning* construction.

Let P be a pair of pants; i.e. a hyperbolic surface with three boundary components. Let $f : P \to M$ be a relative homotopy class of map sending the three boundary components by maps of nonzero degree to three (not necessarily simple) geodesics in M. The class of f determines a homomorphism from $\pi_1(P)$ to $\pi_1(M)$ up to conjugacy.

We give P a hyperbolic structure, and let Δ be a geodesic triangle in P with one vertex on each boundary component. As we move the vertices around on ∂P, the geodesic triangle deforms continuously. *Spinning* Δ involves dragging the vertices around and around the components of ∂P. The sides of Δ get longer and longer, and accumulate on ∂P. The Hausdorff limit of $\partial\Delta$ is a geodesic lamination L in P

with three infinite leaves spiraling around ∂P, and whose complement consists of two (open) ideal triangles. See Figure 3.1.

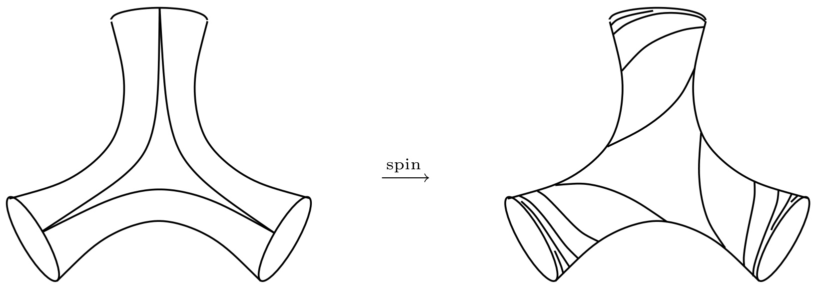


FIGURE 3.1. Spinning produces an ideal triangulation of a pair of pants

We can build a pleated surface $\overline{f} : P \to M$ in the homotopy class of f with pleating locus contained in L as follows. First, $\overline{f}$ takes components of ∂P to the unique closed geodesics in the homotopy class of $f(\partial P)$. For each infinite geodesic $l \in L$, the ends of $f(l)$ spiral around $f(\partial P)$. Except in degenerate cases, the image of $f(l)$ is a quasigeodesic which can be straightened to a unique geodesic $\overline{l}$ which spirals around two components of $\overline{f}(\partial P)$. This defines the map $\overline{f}$ on L. Each component of $P - L$ is an ideal triangle, and we define $\overline{f}$ on each such triangle Δ to be the unique totally geodesic map which extends $\overline{f}$ (after possibly reparameterizing by a translation on each edge) on $\partial\Delta \subset L$.

REMARK 3.5. If M has parabolic elements, the construction in Example 3.4 must be modified very slightly.

Suppose $f : P \to M$ takes some boundary component ∂_0 of P to a free homotopy class in M corresponding to a parabolic conjugacy class α in $\pi_1(M)$. After lifting $\tilde{f} : \tilde{P} \to \tilde{M}$, each conjugate of α fixes a unique point in the sphere at infinity S^2_∞. If Δ is a triangle with a vertex v on ∂_0, and $\tilde{\Delta}$ is a lift of Δ to $\tilde{P}$, straighten $\tilde{f}$ on $\tilde{v}$ by sending this vertex to the unique fixed point of the corresponding conjugate of α. The rest of the construction is as before.

LEMMA 3.6 (Thurston, § 8.10 [**198**]). *A map $f : P \to M$ from a pair of pants into a hyperbolic manifold M can be straightened to a pleated surface unless it factors through a map to a circle.*

PROOF. The map $f : P \to M$ determines an equivariant map $\tilde{P} \to \mathbb{H}^n$ from the universal cover $\tilde{P}$ of P. A lift of the triangle Δ has vertices on three distinct edges e_1, e_2, e_3 of $\tilde{P}$. Spinning drags the vertices of Δ to endpoints of the e_i, so f can be straightened on Δ providing the endpoints of the e_i are distinct for different i. If $\alpha, \beta \in \pi_1(M)$ don't commute, their axes have disjoint endpoints at infinity. Commuting elements in a closed hyperbolic manifold group generate a cyclic group. So the straightening can be achieved if and only if the image of $\pi_1(P)$ in $\pi_1(M)$ does not factor through a cyclic group. □

Using this lemma, we show that a map $f : S \to M$ either admits an obvious simplification which reduces the genus, or has a pleated representative.

LEMMA 3.7. *Let M be a hyperbolic manifold, and let a be a nontrivial conjugacy class in $\pi_1(M)$. Let S be a compact oriented surface with exactly one boundary* *component. If $f : S \to M$ is a map sending the class of ∂S in $\pi_1(S)$ to a then there is another map $f' : S' \to M$ where the genus of S' is no more than that of S, and where f' sends the class of $\partial S'$ in $\pi_1(S')$ to a, which is homotopic to a pleated representative.*

PROOF. We decompose S into subsurfaces $S = S_1 \cup S_2 \cup \cdots \cup S_g$ where each S_i is a twice-punctured torus for $i < g$, and S_g is a once-punctured torus. For each i denote the two boundary components of S_i by $\gamma_i^\pm$ where γ_i^+ is glued to γ_{i+1}^- in S, and γ_1^- maps to a by f.

If any $\gamma_i^\pm$ maps by f to an inessential loop in M, we can compress f, S along the image of this curve, sewing in two disks, to produce a map $f' : S' \to M$ where S' is of smaller genus than S, and for which $f'(\partial S')$ is in the class of a. After finitely many compressions of this kind, we assume that every $\gamma_i^\pm$ maps by f to an essential loop in M.

For each $i < g$, let $\alpha_i^\pm, \beta_i$ be embedded essential loops in S_i as in Figure 3.2:

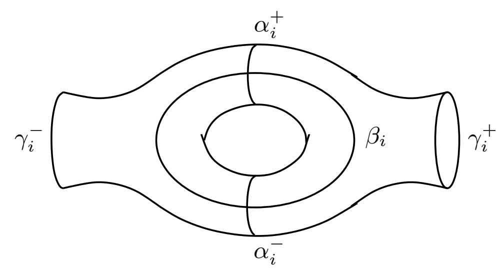


FIGURE 3.2. The curves $\alpha_i^\pm$ and β in S_i

The loops α_i^+, β_i intersect in one point p_i^+. Their images under f define elements a^+, b of $\pi_1(M)$ based at $f(p_i^+)$. If $[a^+, b] = \text{id}$ then we can compress S, cutting out a neighborhood of $\alpha_i^+ \cup \beta_i$ in S_i, and sewing in a disk, thereby reducing the genus of S. So without loss of generality, we assume the elements a^+, b do not commute.

The curves α_i^- and β_i intersect at a different point. By sliding this point half way around β_i and mapping by f, we obtain an element $a^- \in \pi_1(M)$ based at $f(p_i^+)$ if we slide in one direction, and ba^-b^{-1} if we slide in the other direction. Also without loss of generality, we assume a^-, b do not commute.

Let P_i^-, P_i^+ be the two pairs of pants obtained from S_i by cutting along $\alpha_i^+ \cup \alpha_i^-$. Suppose $f : P_i^- \to M$ factors up to homotopy through a map to a circle. This happens if a^+, a^- as above generate a cyclic subgroup of $\pi_1(M)$. In this case, we replace $\alpha_i^\pm$ by their images under a Dehn twist around β_i (see Example 3.59 for a definition). At the level of π_1, this replaces a^+, a^- by $a^+b, b^{-1}a^-$, and defines a new pair of pants decomposition in which $f : P_i^- \to M$ does not factor up to homotopy through a map to a circle. Of course, now $f : P_i^+ \to M$ might factor through a circle, in which case we do another Dehn twist, replacing the original a^+, a^- by $a^+b^2, b^{-2}a^-$. In this way we obtain a decomposition of S_i into two pairs of pants such that the restriction of f to either does not factor up to homotopy through a map to a circle. By Lemma 3.7 f can be replaced by a pleated representative on each such pair of pants, and we are done.

The construction of a pleated representative on S_g is similar but simpler, with $\alpha_i^\pm$ being replaced by a single α_g. □

REMARK 3.8. Other surfaces which perform a similar function include *harmonic* (maps of) surfaces and *minimal surfaces*. The use of one kind of surface or the other is often a matter of taste. One technical advantage of pleated surfaces is that they generalize in some sense to arbitrary δ-hyperbolic groups; see Definition 3.39.

3.2. Spectral Gap Theorem

Let M be a closed hyperbolic manifold. There is a natural bijection between the set of conjugacy classes in $\pi_1(M)$ and the set of closed geodesics in M. It is a fundamental fact that the function

$$\{\text{closed geodesics}\} \xrightarrow{\text{length}} \mathbb{R}$$

which assigns to a closed geodesic its length, is *proper*; i.e. there are only finitely many closed geodesics with length bounded above by any constant. By contrast, if $G = \pi_1(M)$, the function

$$\{\text{conjugacy classes in } [G, G]\} \xrightarrow{\text{scl}} \mathbb{R}$$

which assigns to a (homologically trivial) conjugacy class its stable commutator length, is *not* proper: i.e. there are always infinitely many distinct conjugacy classes with uniformly bounded stable commutator length. However, some vestige of properness holds in this context. If the stable commutator length of a conjugacy class is *sufficiently* small, the length of the corresponding geodesic must also be (comparably) small. This implies that at least for sufficiently small ϵ, the preimage $\text{scl}^{-1}([0, \epsilon])$ is finite. One can define δ_∞ to be the supremum of the set of ϵ with this property; it turns out that there is a universal estimate $\frac{1}{12} \le \delta_\infty \le \frac{1}{2}$.

3.2.1. Length inequality. We now show that in a hyperbolic manifold group, a conjugacy class with sufficiently small stable commutator length is represented by an arbitrarily short geodesic. The material in this section is largely drawn from § 6 of [**42**].

THEOREM 3.9 (Length inequality). *For every dimension m and any $\epsilon > 0$ there is a positive constant $\delta(\epsilon, m)$ such that if M is a complete hyperbolic m-manifold, and a is a conjugacy class in $\pi_1(M)$ with $\text{scl}(a) \le \delta(\epsilon, m)$ then if a is represented by a geodesic γ, we have*

$$\text{length}(\gamma) \le \epsilon$$

PROOF. Let S be a surface of genus g with one boundary component, and $f : S \to M$ a map wrapping ∂S homotopically n times around γ. By Lemma 3.7, after possibly reducing the genus of S if necessary, we can assume without loss of generality that f, S is a pleated surface. This determines a hyperbolic structure on S with geodesic boundary for which the map f is an isometry on paths. In particular, $\text{length}(\partial S) = n \cdot \text{length}(\gamma)$ and $\text{area}(S) = -2\pi\chi(S) = (4g - 2)\pi$ by Gauss–Bonnet.

Choose ϵ which is small compared to the 2-dimensional Margulis constant $\epsilon(2)$. We defer the precise choice of ϵ for the moment. Consider the thick-thin decomposition of S with respect to 2ϵ in the sense of § 3.1.1. More precisely, let DS denote the double of S (which is a closed hyperbolic surface), Let DS_{thick} and DS_{thin} denote the subsets of DS where the injectivity radius is $\ge 2\epsilon$ and $< 2\epsilon$ respectively, and define S_{thick} and S_{thin} to be equal to $DS_{\text{thick}} \cap S$ and $DS_{\text{thin}} \cap S$ respectively.

The set S_{thin} is a union of open embedded annuli around very short simple geodesics, together with a union of open embedded rectangles which run between pairs of segments of ∂S which are distance $< \epsilon$ apart at every point. Each rectangle doubles to an annulus in DS_{thin}. If there are s annuli and r rectangles in S_{thin}, then there are $2s + r$ annuli in DS_{thin}. Components of DS_{thin} are disjoint and pairwise non-isotopic. Any maximal collection of disjoint pairwise non-isotopic simple closed curves in a closed orientable surface of negative Euler characteristic must decompose the surface into pairs of pants. Since the genus of DS is $2g$, we estimate $2s+r \le -\frac{3}{2}\chi(DS) = 6g-3$. Hence r, the number of rectangle components of S_{thin}, is at most $6g - 3$.

By abuse of notation, we add to S_{thick} the annulus components of S_{thin} (if any), so that S_{thin} consists exactly of the set of thin rectangles running between pairs of arcs in ∂S. With this new definition, a point $p \in \partial S$ is in S_{thick} if and only if the length of an essential arc in S from p to ∂S is at least ϵ. In particular, the $\epsilon/2$ neighborhood of $\partial S \cap S_{\text{thick}}$ is *embedded*, and there is an estimate

$$(4g-2)\pi = \text{area}(S) \ge \text{area}(S_{\text{thick}}) \ge \frac{\epsilon}{2}\text{length}(\partial S \cap S_{\text{thick}})$$

Since there are at most $6g - 3$ components of S_{thin}, and each component intersects ∂S in two arcs, there are at most $12g - 6$ components of $\partial S \cap S_{\text{thin}}$. But

$$\text{length}(\partial S \cap S_{\text{thin}}) = \text{length}(\partial S) - \text{length}(\partial S \cap S_{\text{thick}}) \ge n \cdot \text{length}(\gamma) - (8g-4)\frac{\pi}{\epsilon}$$

where we used $\text{length}(\partial S) = n \cdot \text{length}(\gamma)$ and the previous inequality. It follows that there is at least one arc σ of $\partial S \cap S_{\text{thin}}$ satisfying

$$\text{length}(\sigma) \ge \frac{n \cdot \text{length}(\gamma) - (8g-4)\pi/\epsilon}{12g-6} = \frac{n \cdot \text{length}(\gamma)}{12g-6} - \frac{2\pi}{3\epsilon}$$

Hence S_{thin} contains a component R which is a rectangular strip of thickness $\le \epsilon$ with σ on one side. We denote the side opposite to σ by σ'. We call σ and σ' the *long* sides of R. Because S is oriented, the orientations on opposite sides of R are "anti-aligned". We lift R to the universal cover $\mathbb{H}^n$, and by abuse of notation refer to the lifted rectangle as R. The sides σ, σ' of R are contained in geodesics l, l' that cover γ. Without loss of generality, we can suppose that l is an axis for a, and l' is an axis for bab^{-1} where $b(l) = l'$. Moreover, the action of a on l and a' on l' move points is (nearly) opposite directions.

Let p be the midpoint of σ, and let q be a point on the opposite side of R with $d(p, q) < \epsilon$. Suppose further that

$$\text{length}(\sigma) = \text{length}(\sigma') > 2 \cdot \text{length}(\gamma) + 4\epsilon$$

It follows that $bab^{-1}(q) \in \sigma'$ and there is $r \in \sigma$ with $d(bab^{-1}(q), r) \le \epsilon$ and therefore $d(bab^{-1}(p), r) \le 2\epsilon$. Since $d(q, bab^{-1}(q)) = \text{length}(\gamma)$,

$$|d(p,r) - \text{length}(\gamma)| \le 2\epsilon$$

and therefore $d(p, a(r)) \le 2\epsilon$ and we can estimate

$$d(p, abab^{-1}(p)) \le 4\epsilon$$

Similarly we estimate $d(p, bab^{-1}a(p)) \le 4\epsilon$. See Figure 3.3. In the figure, the axes l and l' are both roughly vertical. The element a translates points roughly downwards along l, and bab^{-1} translates points roughly upwards along l'.

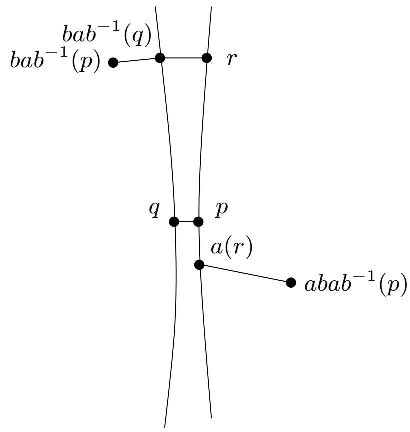


FIGURE 3.3. The composition $abab^{-1}$ translates the midpoint p a small distance

If we choose 4ϵ less than an m-dimensional Margulis constant $\epsilon(m)$ then $abab^{-1}$ and $bab^{-1}a$ must commute. There are two possibilities, which break up into subcases.

CASE ($abab^{-1}$ and $bab^{-1}a$ are hyperbolic with the same axis). In this case, since they are conjugate, they are either equal or inverse.

SUBCASE ($abab^{-1} = bab^{-1}a$). In this case a and bab^{-1} commute, and since they are conjugate, they are equal or inverse. But a and bab^{-1} translate their respective axes in almost opposite directions, so they cannot be equal; hence we must have $bab^{-1} = a^{-1}$ and therefore b has order 2, which is impossible in a hyperbolic manifold group.

SUBCASE ($abab^{-1} = a^{-1}ba^{-1}b^{-1}$). In this case $a^2 = ba^{-2}b^{-1}$ and therefore b has order 2, which is impossible as we already remarked.

CASE ($abab^{-1}$ and $bab^{-1}a$ parabolic with the same fixed point). $z \in S^{n-1}_\infty$. In this case, $a^{-1}(abab^{-1})a$ is parabolic with fixed point $a^{-1}(z)$. But $a^{-1}(abab^{-1})a = bab^{-1}a$ which has fixed point z, so $a^{-1}(z) = z$. Since a translates along an axis, it is hyperbolic, and we have obtained a hyperbolic and a parabolic element in $\pi_1(M)$ with a common fixed point at infinity. This is well-known to violate discreteness, see for instance Maskit [**147**], p. 19 for details.

In every case we obtain a contradiction, and therefore we must have

$$2 \cdot \text{length}(\gamma) + 4\epsilon \geq \text{length}(\sigma)$$

Putting this together with our earlier inequality, we obtain

$$2 \cdot \text{length}(\gamma) + 4\epsilon \geq \frac{n \cdot \text{length}(\gamma)}{12g - 6} - \frac{2\pi}{3\epsilon}$$

Rearranging this gives

$$\text{length}(\gamma) \cdot \left(\frac{n}{12g - 6} - 2 \right) \leq 4\epsilon + \frac{2\pi}{3\epsilon}$$

The right hand side is a constant which depends only on the size of a Margulis constant in dimension n. If scl is very small, we can make n/g very large, and therefore obtain an upper bound on length(γ) which goes to 0 as scl $\to 0$ as claimed. □

REMARK 3.10. Note that when $m < n$ a hyperbolic m-manifold group is also a hyperbolic n-manifold group, and therefore $\delta(\epsilon, m) \geq \delta(\epsilon, n)$. In § 3.3 we will see that for small ϵ there are estimates

$$\delta(\epsilon, 3) = O(\epsilon^{1/2})$$

and

$$\delta(\epsilon, 3) \geq \delta(\epsilon, n) \geq O(\epsilon^{(n-1)/(n+1)})$$

in any fixed dimension n.

On the other hand, the dependence of δ on ϵ is not proper. In particular, as $\epsilon \to \infty$, the constant $\delta(\epsilon, n)$ is bounded above by some finite bound, independent of dimension n. This universal upper bound should be thought of as a kind of homological Margulis constant. In the next subsection, we will give an explicit estimate for this constant.

3.2.2. Spectral Gap.

THEOREM 3.11 (Spectral Gap Theorem). *Let M be a closed hyperbolic manifold, of any dimension ≥ 2. Let $\delta_\infty(M)$ be the first accumulation point for stable commutator length on conjugacy classes in $\pi_1(M)$. That is, $\delta_\infty(M)$ is the smallest number such that for any $\delta < \delta_\infty(M)$, there are only finitely many conjugacy classes a in $\pi_1(M)$ with $\mathrm{scl}(a) \leq \delta$. Then*

$$\frac{1}{12} \leq \delta_\infty(M) \leq \frac{1}{2}$$

PROOF. We use the same setup and notation as in the proof of Theorem 3.9. Since M is a closed hyperbolic manifold, there are only finitely many conjugacy classes represented by geodesics shorter than any given length. So we suppose a is a conjugacy class represented by a geodesic γ which is "sufficiently long" (in a sense to be made precise in a moment). We choose ϵ and find a segment σ, as in the proof of Theorem 3.9, and suppose we have

$$\text{length}(\gamma) + 4\epsilon < \text{length}(\sigma)$$

(note the missing factor of 2). We choose p to be one of the endpoints of σ, so that

$$d(p, abab^{-1}(p)) \leq 4\epsilon$$

Since M is fixed, there is some ϵ such that 4ϵ is smaller than the translation length of any nontrivial element in $\pi_1(M)$. Hence $abab^{-1} = \mathrm{id}$. But this means $bab^{-1} = a^{-1}$, and b has order 2, which is impossible in a manifold group.

Contrapositively, this means that we must have

$$\text{length}(\gamma) + 4\epsilon \geq \text{length}(\sigma)$$

and therefore, just as in the proof of Theorem 3.9, we obtain

$$\text{length}(\gamma) \cdot \left(\frac{n}{12g - 6} - 1 \right) \leq 4\epsilon + \frac{2\pi}{3\epsilon}$$

In contrast to the case of Theorem 3.9, the right hand side definitely depends on the manifold M. Nevertheless, for fixed M, it is a constant, and we see that for γ

sufficiently long, g/n cannot be much smaller than $1/12$. This establishes the lower bound in the theorem.

We now establish the upper bound. M is a closed hyperbolic manifold, and therefore $\pi_1(M)$ contains many nonabelian free groups. In fact, if a, b are arbitrary noncommuting elements of $\pi_1(M)$, sufficiently high powers of a and b generate a free group, by the ping-pong lemma. This copy of F_2 is quasi-isometrically embedded, and by passing to a subgroup, one obtains quasi-isometrically embedded copies of free groups of any rank.

For each n, the element $[a^n, b^n]$ is in the commutator subgroup. In fact, it is a commutator, and therefore satisfies $\text{scl}([a^n, b^n]) \le 1/2$. In a free group, the words $x^n y^n x^{-n} y^{-n}$ are cyclically reduced of length $4n$. Since the embedding is quasi-isometric, the geodesic representatives of $[a^n, b^n]$ have length which goes to infinity linearly in n. It follows that these elements fall into infinitely many conjugacy classes, and the upper bound is established. □

REMARK 3.12. From the method of proof one sees for sufficiently long γ that if no translate l' of l is ϵ-close and anti-aligned with l along segments σ, σ' whose length is at least $(\lambda + \epsilon) \cdot \text{length}(\gamma)$ then $\text{scl}(a) \ge \frac{1}{12\lambda}$.

For example, in a free group, a cyclically reduced word w and a conjugate of its inverse cannot share a subword of length longer than $\frac{1}{2}\text{length}(w)$. This leads to an estimate $\text{scl}(a) \ge 1/6$ in a free group, which is not yet optimal, but is still an improvement (a sharp bound $\text{scl}(a) \ge 1/2$ in a free group will be established in Theorem 4.111).

An estimate on the size of anti-aligned translates is essentially a kind of macroscopic small cancellation property. One can give an alternative proof of Theorem 3.11 along these lines using generalized small cancellation theory (see [**65**] for more details). For certain groups, ordinary small cancellation theory can be applied, leading to sharp results; we will discuss this approach in § 4.3.

3.3. Examples

3.3.1. Hyperbolic Dehn surgery.

We elaborate on the discussion in § 3.1.2.

LEMMA 3.13. *Let M be a hyperbolic 3-manifold, and let γ be a geodesic loop which is the core of an embedded solid torus of radius T. Then there is a 1-form α supported in the tube of radius T about γ, with $\int_\gamma \alpha = \text{length}(\gamma)\sinh(T)$ and $\|d\alpha\| \le 1 + 1/(T - \epsilon)$ for any $\epsilon > 0$.*

PROOF. Let S be the solid torus of radius T about γ. On S, let $r : S \to \mathbb{R}$ be the function which measures distance to γ. Denote radial projection to γ by

$$p : S \to \gamma$$

Parameterize γ by θ, so that $d\theta$ is the length form on γ, and $\int_\gamma d\theta = \text{length}(\gamma)$. Pulling back by p extends θ and $d\theta$ to all of S. We define

$$\alpha = d\theta \cdot (\sinh(T) - \sinh(r))$$

on S, and extend it by 0 outside S. Notice that

$$\|d\theta\| = 1/\cosh(r)$$

on S. By direct calculation, $d\alpha = \cosh(r) d\theta \wedge dr$ on S, so $\|d\alpha\| = 1$ at every point of S.

The form α is not smooth along ∂S, but it is Lipschitz. Let $\beta_\epsilon(r)$ be a C^∞ function on $[0, T]$ taking the value 1 in a neighborhood of 0 and the value 0 in a neighborhood of T, and with $|\beta'_\epsilon| < 1/(T-\epsilon)$ throughout, for some small ϵ. The product $\alpha_\epsilon := \beta_\epsilon(r)\alpha$ is C^∞ and satisfies

$$d\alpha_\epsilon = d\theta \wedge dr(\beta_\epsilon(r)\cosh(r) + \beta'_\epsilon(r)\sinh(r))$$

so $\|d\alpha_\epsilon\| \le 1 + 1/(T-\epsilon)$. □

As in § 2.3.1 there is a de Rham quasimorphism q_α associated to α by integration over based geodesic representatives of elements, after choosing a basepoint. The homogenization of q_α is obtained by integrating α over free geodesic loops. A limit of such quasimorphisms as $\epsilon \to 0$ has defect at most $2\pi(T+1)/T$ by Lemma 2.58.

In order for Lemma 3.13 to be useful, we need a good estimate of T in terms of length(γ).

LEMMA 3.14 (Hodgson–Kerckhoff, p. 403 [**111**]). *Let S be a Margulis tube in a hyperbolic 3-manifold. Let T be the radius of S and* length(γ) *the length of the core geodesic. Then there is an estimate*

$$\text{length}(\gamma) \ge 0.5404 \frac{\tanh(T)}{\cosh(2T)}$$

Note for γ sufficiently small this implies $e^T \ge 1.03\,\text{length}^{-1/2}(\gamma)$.

REMARK 3.15. In any dimension n a much cruder argument due to Reznikov [**177**] shows that for sufficiently small γ there is a constant C_n such that $e^T \ge C_n\,\text{length}^{-2/(n+1)}(\gamma)$.

Now fix M, a 1-cusped hyperbolic 3-manifold. Fix generators m, l for $H_1(\partial M)$ for which l generates the kernel of $H_1(\partial M;\mathbb{Q}) \to H_1(M;\mathbb{Q})$. Let $M_{p/q}$ denote the result of p/q Dehn surgery on M in these co-ordinates, and let $\gamma(p/q)$, or just γ for short, denote the core geodesic of the filled solid torus.

THEOREM 3.16. *Let $M_{p/q}$ be the result of p/q surgery on M. Suppose $M_{p/q}$ is hyperbolic. When the core geodesic γ is contained in a Margulis tube of radius at least T then*

$$\text{length}(\gamma) \le \left(\frac{7.986\pi\,\text{scl}(l)(T+1)}{Tp}\right)^2$$

PROOF. By Lemma 3.13 there is a homogeneous quasimorphism q_α on $\pi_1(M_{p/q})$ with defect at most 2π, and satisfying

$$q_\alpha(\gamma) \ge \text{length}(\gamma)\sinh(T)\frac{T}{T+1}$$

On the other hand, the conjugacy class of γ^p contains the image of l under the surjective homomorphism $\pi_1(M_\gamma) \to \pi_1(M_{p/q})$ induced by Dehn surgery, so by the easy direction of Bavard's Duality Theorem 2.70, we estimate

$$\frac{q_\alpha(\gamma)}{4\pi} \le \text{scl}(\gamma) \le \frac{\text{scl}(l)}{p}$$

Using the estimate from Lemma 3.14, a straightforward calculation gives the desired conclusion. □

Neumann–Zagier [**161**] introduce the following quadratic form Q:

$$Q(p,q) = \frac{(\text{length of } pm + ql)^2}{\text{area}(\partial S)}$$

Here ∂S is the horotorus boundary of the cusp of M, and $pm + ql$ is a straight curve on the horotorus (in the intrinsic Euclidean metric) representing $m^p l^q$. Equivalently, if we scale the Euclidean cusp to have area 1, the form just becomes $Q(p,q) = \text{length}^2(pm + ql)$.

LEMMA 3.17 (Neumann–Zagier, Prop. 4.3 [**161**]). *With notation as above, in the manifold $M_{p/q}$ there is an estimate*

$$\text{length}(\gamma) = 2\pi Q(p,q)^{-1} + O\left(\frac{1}{p^4 + q^4}\right)$$

In particular, for q fixed, there is an estimate

$$\lim_{p\to\infty} (pm)^2 \text{length}(\gamma)/2\pi = 1$$

where m is the length of the meridian in the Euclidean cusp, normalized to have area 1.

REMARK 3.18. We see from Lemma 3.17 that the estimates obtained in Theorem 3.16 are sharp, up to an order of magnitude. Together with Remark 3.15, this justifies the claims made in Remark 3.10.

THEOREM 3.19. *Let M be a 1-cusped hyperbolic manifold, with notation as above. Normalize the Euclidean structure on the cusp ∂S to have area 1, and let m be the length of the shortest curve on ∂S which is homologically essential in M. If* $\text{length}(m) < 1$ *then*

$$\text{scl}(l) \geq \frac{1}{4\pi\,\text{length}(m)^2}$$

PROOF. For brevity, we denote (normalized) length(m) by m. We expand S to a maximal horotorus. For a maximal horotorus, every essential slope on ∂S has length at least 1, by Jørgensen's inequality [**147**]. It follows that if $m < 1$, then $\text{area}(\partial S) \geq 1/m^2$. Under p/q surgery for very large p, the area of the boundary of a maximal embedded tube around γ is almost equal to that of area(∂S). The boundary of such a tube is intrinsically Euclidean in its induced metric, and is isometric to a torus obtained from a product annulus by gluing the two end components with a twist. The boundary components of the annulus have have length equal to the circumference of a circle in the hyperbolic plane of radius T, which is $2\pi\sinh(T)$. By elementary hyperbolic trigonometry, the height of the annulus is equal to $\text{length}(\gamma)\cosh(T)$. Hence the area of the boundary of the tube is $2\pi\,\text{length}(\gamma)\sinh(T)\cosh(T)$.

So we can estimate

$$\text{area}(\partial S) = \lim_{p\to\infty} 2\pi\,\text{length}(\gamma)\sinh(T)\cosh(T)$$

and therefore

$$e^T \geq \frac{\sqrt{2}}{m\sqrt{\pi}}\,\text{length}^{-1/2}(\gamma)$$

Using this estimate in the place of Lemma 3.14 in Theorem 3.16, and applying Lemma 3.17, we obtain

$$\frac{2\pi}{(pm)^2} = \lim_{p\to\infty} \text{length}(\gamma) \leq \left(\frac{4\text{scl}(l) m\pi\sqrt{2\pi}}{p}\right)^2$$

and therefore

$$\mathrm{scl}(l) \ge \frac{1}{4\pi m^2}$$

as claimed □

In other words, one can estimate $\mathrm{scl}(l)$ from below from the geometry of the cusp.

3.3.2. Manifolds with small δ_∞. Note that the proof of Theorem 3.11 actually shows that if M is any closed hyperbolic manifold, and a is a conjugacy class in $\pi_1(M)$ represented by a geodesic γ, then if $\mathrm{length}(\gamma)$ is sufficiently long, $\mathrm{scl}(a) \ge \delta$ for any $\delta < 1/12$.

Example 3.20. Let S be a closed nonorientable surface with $\chi(S) = -1$. A presentation for $\pi_1(S)$ is

$$\langle a, b, c \mid [a, b] = c^2 \rangle$$

so the conjugacy class of c satisfies $\mathrm{scl}(c) \le 1/4$. On the other hand, for a suitable choice of hyperbolic structure on S, the geodesic in the free homotopy class of c can be arbitrarily long.

Question 3.21. *What are the optimal constants in Theorem 3.11?*

We will see in § 4.3.4 that the upper bound of 1/2 is sharp, and is realized in free and orientable surface groups.

Example 3.22. For any group G and any elements $a, b \in G$ the element $[a, b]$ satisfies $\mathrm{scl}([a, b]) \le 1/2$. Moreover, by Proposition 2.104, if a and b do not generate a free rank 2 subgroup of G, we must have $\mathrm{scl}([a, b]) < 1/2$.

However, a theorem of Delzant [**64**] shows that in any word-hyperbolic group G (see § 3.4 for a definition) there are only finitely many conjugacy classes of non-free 2-generator subgroups. Note that this class of groups includes fundamental groups of closed hyperbolic manifolds of any dimension. Therefore only finitely many conjugacy classes of elements $[a, b]$ with $\mathrm{scl}([a, b]) < 1/2$ can be constructed in a fixed hyperbolic group G this way.

3.3.3. Complex length. If M is a closed hyperbolic 3-manifold, a conjugacy class $a \in \pi_1(M)$ determines a geodesic γ which has a complex length, denoted $\mathrm{length}_{\mathbb{C}}(\gamma)$, defined as follows. The hyperbolic structure on M determines a representation $\rho : \pi_1(M) \to \mathrm{PSL}(2, \mathbb{C})$. The trace $\mathrm{tr}(a)$ is well-defined up to multiplication by ± 1. We set

$$\mathrm{length}_{\mathbb{C}}(\gamma) = \cosh^{-1}(\mathrm{tr}(a)/2)$$

which is well-defined up to integral multiples of $2\pi i$. The real part of $\mathrm{length}_{\mathbb{C}}$ is the usual length of γ, and the imaginary part is the angle of rotation on the normal bundle $\nu\gamma$ to γ induced by parallel transport around γ.

If γ is trivial in $H_1(M; \mathbb{Q})$ there is a slightly different $\mathbb{C}$-valued complex length, denoted $\mathrm{length}_H(\gamma)$, and defined as follows. Let $N(\gamma)$ be an open solid torus neighborhood of γ, and let T be the torus boundary of $M - N(\gamma)$. Let l be the slope on T which generates the kernel of the map $H_1(T; \mathbb{Q}) \to H_1(M - N(\gamma); \mathbb{Q})$. If M is obtained from $M - N(\gamma)$ by p/q filling with respect to some basis (q is arbitrary, depending on a choice of meridian m, but p is well-defined) then l determines a framing of $\nu\gamma^p$, the normal bundle of the p-fold cover of γ. This framing determines the imaginary part of $\mathrm{length}_H(\gamma)$; in words, minus the imaginary part is the angle that the framing l twists relative to parallel transport around γ. Note that the imaginary part of $\text{length}_H(\gamma)$ and the imaginary part of $\text{length}_{\mathbb{C}}(\gamma)$ will only agree up to integral multiples of $2\pi i/p$. Note that for a fixed choice of meridian m one can estimate

$$\text{imaginary part of length}_H(\gamma) = \text{some function of } p, q + O(1)$$

In terms of differential forms: near γ we can define cylindrical co-ordinates θ, ϕ, r where θ parameterizes the length along γ and r is distance to γ, as in Lemma 3.13, and where ϕ is the angular co-ordinate, taking values (locally) in $\mathbb{R}/2\pi\mathbb{Z}$. The co-ordinate ϕ is not globally well-defined unless $\text{length}_{\mathbb{C}}(\gamma)$ has imaginary part which is a multiple of $2\pi i$, but the forms $d\phi$ and $d\theta$ are well-defined. With respect to this co-ordinate system,

$$\text{length}_H(\gamma) = \frac{1}{p}\int_l d\theta + i d\phi$$

In analogy to the construction in Lemma 3.13, define $\beta = d\phi(\cosh(T) - \cosh(r))$ and observe that

$$\|d(\alpha + i\beta)\| = 1$$

on $T - \gamma$. One must be careful, since $d\phi$ does not extend over γ. Nevertheless, if S is a surface in $M - N(\gamma)$ of genus g whose boundary wraps m times around l, we can represent S by a pleated surface in M. For sufficiently large p or q the length (in the usual sense) of γ will be very short, and any surface S which intersects γ transversely will have area at least e^T. In particular, for all but finitely many surgeries, a pleated representative of S in M is disjoint from γ, and we obtain an estimate of the form

$$|\text{length}_H(\gamma)| \le \text{some function of } \text{scl}(l), p, q$$

valid for large p or q, which refines the inequality in Theorem 3.16.

3.4. Hyperbolic groups

We would like to generalize Theorem 3.9 and Theorem 3.11 beyond fundamental groups of hyperbolic manifolds to more general (word) hyperbolic groups. There are two essential ingredients in the proof of these theorems:

(1) the existence of a pleated surface representative in each homotopy class
(2) the existence of a Margulis constant in each dimension n

In fact, the proof of Theorem 3.11 only uses the existence of a Margulis constant in dimension 2, and the fact that in a given closed hyperbolic manifold there is a uniform positive lower bound on the translation length of any element.

We will see that both of these ingredients have acceptable generalizations to the context of hyperbolic groups, and therefore we obtain generalizations of these theorems with similar (geometric) proofs.

Alternatively, these theorems can be proved by explicitly constructing quasimorphisms with suitable properties and appealing to (the easy direction of) Bavard duality. The construction of quasimorphisms on hyperbolic groups extends to groups acting (weakly properly discontinuously) on hyperbolic spaces, such as mapping class groups, groups acting on trees, $\text{Out}(F_n)$ and so on, as we shall see in subsequent sections.

Where it pertains to quasimorphisms and stable commutator length, the material in the remainder of this chapter draws substantially on [**13, 12, 49**]. We also appeal to [**98, 24, 156**] for facts about hyperbolic spaces and groups, and [**21, 148**] for facts about the geometry of the curve complex.

3.4.1. Definitions and basic properties. Let G be a group with a finite symmetric generating set A. Let $C_A(G)$ be the Cayley graph of G with respect to A. In other words, $C_A(G)$ is the graph with one vertex for each element of G, and one edge from vertices g to g' for each pair of elements $g, g' \in G$ and each $a \in A$ for which $g' = ga$. We make $C_A(G)$ into a path metric space by declaring that the length of every edge is 1. The left action of G on itself extends to a simplicial (and therefore isometric) action of G on $C_A(G)$. Providing A contains no elements of order 2, the action is free and cocompact, with quotient a wedge of $|A|$ circles.

Definition 3.23. A path metric space is *δ-hyperbolic* for some $\delta \geq 0$ if for every geodesic triangle abc, every point in the edge ab is contained in the union of the δ-neighborhoods of the other two edges:

$$ab \subset N_\delta(bc) \cup N_\delta(ca)$$

A group G with a finite symmetric generating set A is *δ-hyperbolic* if $C_A(G)$ is δ-hyperbolic as a path metric space.

G is *word-hyperbolic* (or simply *hyperbolic*) if there is a $\delta \geq 0$ and a finite symmetric generating set A for which $C_A(G)$ is δ-hyperbolic.

Remark 3.24. Note that our definition of a δ-hyperbolic space requires it to be a path metric space; other definitions (e.g. in terms of the Gromov product) do not require this.

Example 3.25. Finitely generated free groups are hyperbolic. Fundamental groups of compact surfaces with $\chi < 0$ are hyperbolic.

Example 3.26. Let M be a closed Riemannian manifold with sectional curvature uniformly bounded above by a negative number. Then $\pi_1(M)$ is hyperbolic.

Example 3.27. A group with a presentation satisfying the small cancellation condition $C(7)$ (see § 4.3) is hyperbolic.

Example 3.28. A group $G = \langle X_m \mid R \rangle$ on a finite generating set X_m with a "random" set of relations R, drawn according to a suitable probability law (see [**163**]) is hyperbolic with probability 1.

In some sense, "most" groups are hyperbolic. On the other hand, many naturally occurring classes of groups (e.g. amenable groups, $\mathrm{SL}(n, \mathbb{Z})$ for $n \geq 3$, fundamental groups of cusped hyperbolic manifolds of dimension at least 3) are not hyperbolic. Nevertheless, hyperbolic groups are central in geometric group theory.

Remark 3.29. If G is δ-hyperbolic with respect to a generating set A, there is an n depending only on δ so that if A_n denotes the set of elements in G of word length at most n, then G is 7-hyperbolic with respect to the generating set A_n. Hence δ may be taken to be some fixed small number at the expense of possibly increasing $|A|$. On the other hand, δ cannot be made arbitrarily small: a graph is 0-hyperbolic if and only if it is a tree. If $C_A(G)$ is a tree, then G is free, and A is a free generating set for G.

We assume the reader is familiar with basic elements of coarse geometry: (k, ϵ)-quasi-isometries, quasigeodesics, etc. We summarize some of the main properties of δ-hyperbolic spaces below (see [**98**] or [**24**] for details):

Theorem 3.30 (Basic properties of hyperbolic spaces). *Let X be a δ-hyperbolic path metric space.*

(1) **Morse Lemma**. *For every k, ϵ there is a universal constant $C(\delta, k, \epsilon)$ such that every (k, ϵ)-quasigeodesic segment with endpoints $p, q \in X$ lies in the C-neighborhood of any geodesic joining p to q.*
(2) **Quasigeodesity is local**. *For every k, ϵ there is a universal constant $C(\delta, k, \epsilon)$ such that every map $\phi : \mathbb{R} \to X$ which restricts on each segment of length C to a (k, ϵ)-quasigeodesic is (globally) $(2k, 2\epsilon)$-quasigeodesic.*
(3) **Ideal boundary**. *There is an* ideal boundary *∂X functorially associated to X, whose points consist of quasigeodesic rays up to the equivalence relation of being a finite Hausdorff distance apart. There is a natural topology on ∂X for which it is metrizable. If X is proper, ∂X is compact. Moreover, any quasi-isometric embedding $X \to Y$ between hyperbolic spaces induces a continuous map $\partial X \to \partial Y$.*

If G is hyperbolic, we denote the ideal boundary of its Cayley graph by ∂G. As a topological space, this does not depend on the choice of a generating set, so we call it the *ideal boundary* (or just the *boundary*) of G. The left action of G on itself induces an action of G on ∂G by homeomorphisms. Every element $g \in G$ is either finite order (i.e. is *elliptic*), or fixes two points $p^{\pm}$ in ∂G with "source-sink" dynamics (i.e is *hyperbolic*). That is, for any $q \in \partial G - p^{\pm}$ and any neighborhood U of p^+, the translate $g^n(q)$ lies in U for all sufficiently large positive n. The point p^+ is called the *attracting fixed point* of g, and p^- is called the *repelling fixed point*. Note that p^- is the attracting fixed point and p^+ the repelling fixed point for g^{-1}.

In fact, hyperbolic groups are completely characterized by the dynamics of their action on the boundary. The following characterization is due to Bowditch.

Theorem 3.31 (Bowditch, [**20**]). *Let M be a perfect metrizable compact Hausdorff space. Let G be a group acting faithfully on M by homeomorphisms. Let M^3 denote the space of* distinct *ordered triples of elements of M; i.e. the open subset of $M \times M \times M$ consisting of triples which are pairwise distinct. If the induced action of G on M^3 is properly discontinuous and cocompact, then G is hyperbolic, and there is a G-equivariant homeomorphism from M to ∂G.*

It is straightforward to show that a hyperbolic group acts on its boundary as in Theorem 3.31 and therefore this theorem gives a complete characterization of hyperbolic groups. If G is hyperbolic and ∂G contains more than two points, Klein's ping-pong argument applied to the action of G on ∂G shows that G contains many (quasi-isometrically embedded) nonabelian free groups of arbitrary finite rank. A hyperbolic group for which ∂G contains at most two points is said to be *elementary*; a group is elementary hyperbolic if and only if it is virtually cyclic.

Definition 3.32. If X is a metric space, and $g \in \text{Isom}(X)$, the *translation length* of g, denoted $\tau(g)$, is the limit

$$\tau(g) = \lim_{n \to \infty} \frac{d_X(p, g^n(p))}{n}$$

where $p \in X$ is arbitrary.

The triangle inequality implies that this limit exists and is independent of p (and is therefore a conjugacy invariant). If X is a path metric space, and g fixes some geodesic l and acts on it as a translation, then $\tau(g) = d_X(q, g(q))$ for any $q \in l$. If G is a word-hyperbolic group and A is a generating set, then for any $g \in G$ the translation length $\tau_A(g)$, or just $\tau(g)$ if A is understood, is the translation length of g thought of as an element of $\text{Isom}(C_A(G))$ under the natural left action of G on itself. Algebraically, $\tau(g) = \lim_{n\to\infty} \|g^n\|_A/n$ where $\|\cdot\|$ denotes word length with respect to the generating set A.

EXAMPLE 3.33. Let G be any group, and let S denote the set of commutators in G. Then the commutator subgroup $[G,G]$ acts on $C_S([G,G])$ by isometries, and for every $g \in [G,G]$ there is an equality $\text{scl}(g) = \tau(g)$.

The following Lemma is an easy consequence of the local finiteness of $C_A(G)$, the fact that quasigeodesity is local, and the Morse Lemma.

LEMMA 3.34 (Axes in hyperbolic Cayley graphs). *Let G be δ-hyperbolic with respect to the generating set A. Then there is a positive constant $C(\delta,|A|)$ such that every $g \in G$ either has finite order, or there is some $n \le C$ such that g^n fixes some bi-infinite geodesic axis l_g and acts on it by translation.*

For a proof, see Theorem 5.1 from [**78**], or [**24**].

COROLLARY 3.35. *Let G be δ-hyperbolic with respect to the generating set A. Then there is a positive constant $C'(\delta,|A|)$ such that every $g \in G$ either has finite order, or satisfies $\tau(g) \ge C'$.*

PROOF. Since $C_A(G)$ is a graph in which every edge has length 1, elements of $\text{Isom}(C_A(G))$ act on $C_A(G)$ simplicially. It follows that if an element $\gamma \in \text{Isom}(C_A(G))$ acts on some geodesic l by translation, then $\tau(\gamma)$ is an integer. Now apply Lemma 3.34. □

REMARK 3.36. The same argument shows that for a fixed hyperbolic group G, there is a constant $n(\delta,|A|)$ so that $\tau(g) \in \frac{1}{n}\mathbb{Z}$ for all $g \in G$.

3.4.2. Mineyev's flow space. The main difference between hyperbolic manifolds and Cayley groups of hyperbolic groups is *synchronous exponential convergence* of asymptotic geodesics. Two asymptotic geodesic rays in the hyperbolic plane have parameterizations by length such that the distance between corresponding points goes to 0 like e^{-t}. In a word-hyperbolic group, asymptotic geodesic rays eventually come within distance δ of each other, but may not get any closer. It is this synchronous exponential convergence which lets one estimate area from topology in hyperbolic surfaces, and it is crucial for our applications.

It is a fundamental insight due originally to Gromov that the geometry of a δ-hyperbolic space becomes much more tractable when one considers as primitive elements not *points*, but (bi-infinite) *geodesics*. Mineyev gave a precise codification of this insight, and constructed a geometric *flow space* associated to a δ-hyperbolic metric space, in which synchronous exponential convergence of asymptotic geodesics is restored.

A bi-infinite geodesic in a δ-hyperbolic space X contains two distinct geodesic rays, which are asymptotic to distinct points in ∂X. Conversely, if X is a *proper* metric space (i.e. the closed balls of any radius are compact) then any two distinct points in ∂X are the endpoints of some infinite geodesic.

We use the abbreviation $\partial^2 X$ to denote the space of ordered pairs of distinct points in ∂X:

$$\partial^2 X = \{(a,b) \in \partial X \times \partial X \text{ for which } a \neq b\}$$

Mineyev's flow space is not quite a metric space but rather a *pseudo-metric space*, i.e. a space together with a non-negative function $d(\cdot,\cdot)$ on pairs of points which satisfies all the axioms of a metric space except that $d(p,q)$ should be strictly positive for distinct points p and q. The reason is that Mineyev's space is a union (in a suitable sense) of *oriented* geodesics. Two geodesics with opposite orientation corresponding to the same (equivalence class of) geodesic in X cannot be distinguished by the distance function. However, there is a natural quotient of Mineyev's flow space in which these distinct oriented geodesics are identified, and the function d descends to a genuine metric on the quotient.

THEOREM 3.37 (Mineyev's flow space [**156**]). *Let X, d_X be a δ-hyperbolic graph whose vertices all have valence $\le n$. Then there exists a pseudo-metric space $\mathcal{F}(X), d^\times$ called the* flow space of X *with the following properties:*

(1) *$\mathcal{F}(X)$ is homeomorphic to $\partial^2 X \times \mathbb{R}$. The factors $(p,q,\cdot)$ under this homeomorphism are called the* flowlines.
(2) *There is an $\mathbb{R}$-action on $\mathcal{F}(X)$ (the* geodesic flow*) which acts as an isometric translation on each flowline $(p,q,\cdot)$.*
(3) *There is a $\mathbb{Z}/2\mathbb{Z}$ action $x \to x^*$ which anti-commutes with the $\mathbb{R}$ action, which satisfies $d^\times(x,x^*) = 0$, and which interchanges the flowlines $(p,q,\cdot)$ and $(q,p,\cdot)$.*
(4) *There is a natural action of* $\text{Isom}(X)$ *on $\mathcal{F}(X)$ by isometries. If $g \in$* $\text{Isom}(X)$ *is hyperbolic with fixed points $p^\pm$ in ∂X then g fixes the flowline $(p^-,p^+,\cdot)$ of $\mathcal{F}(X)$ and acts on it as a translation by a distance which we denote $\tau(g)$. This action of* $\text{Isom}(X)$ *commutes with the $\mathbb{R}$ and $\mathbb{Z}/2\mathbb{Z}$ actions.*
(5) *There are constants $C \ge 0$ and $0 \le \lambda < 1$ such that for all triples $a,b,c \in \partial X$, there is a natural isometric parameterization of the flowlines $(a,c,\cdot),(b,c,\cdot)$ for which there is exponential convergence*
$$d^\times((a,c,t),(b,c,t)) \le C\lambda^t$$
Explicitly, $(a,c,0)$ is the point on $(a,c,\cdot)$ closest to b, and similarly for $(b,c,0)$ and a (as measured by suitable horofunctions).
(6) *If X admits a cocompact isometric action, then up to an additive error, there is an* $\text{Isom}(X)$ *equivariant (k,ϵ) quasi-isometry between $\mathcal{F}(X), d^\times$ and X, d_X.*

Moreover, all constants as above depend only on δ and n.

This theorem conflates several results and constructions in [**156**]. The pseudometric $d^\times$ is defined in § 3.2 and § 8.6 on a slightly larger space which Mineyev calls the *symmetric join*. The flow space, defined in § 13, is a natural subset of this space. The basic properties of the $\mathbb{R}, \mathbb{Z}/2\mathbb{Z}$ and $\text{Isom}(X)$ action are established in § 2. The remaining properties are subsets of Theorem 44 (p. 459) and Theorem 57 (p. 468).

There are a number of subtle details in the statement of this theorem, which require some discussion.

Bullet (3) implies that after quotienting $\mathcal{F}(X)$ by $\mathbb{Z}/2\mathbb{Z}$, the flowlines $(p,q,\cdot)$ and $(q,p,\cdot)$ become identified, and we can speak of *the* (unparameterized) geodesic joining p and q in the quotient which we denote $\overline{\mathcal{F}}(X)$. By abuse of notation, for each pair of distinct points $p,q \in \partial X$, let $(p,q,\cdot)$ denote a particular isometric parameterization of the unique geodesic in $\overline{\mathcal{F}}(X)$ joining p to q.

Bullet (5) is precisely the *synchronous* exponential convergence of flowlines which is achieved in hyperbolic space, but which is not achieved in hyperbolic groups. We refer to the special isometric synchronous parameterizations of asymptotic geodesics in this bullet as *nearest point parameterizations*. Note that nearest point parameterizations also make sense for ideal triangles in hyperbolic space, or hyperbolic space scaled to have constant curvature K for any $K < 0$. We define an *ideal triangle* in $\overline{\mathcal{F}}(X)$ to be the union of three (unparameterized) geodesics joining distinct $a, b, c \in \partial X$ in pairs.

The additive error in Bullet (6) spoiling genuine equivariance is necessary in case $\mathrm{Isom}(X)$ is indiscrete or does not act freely on X. If X is the Cayley graph of a torsion-free hyperbolic group G, then G acts freely on both $\mathcal{F}(X)$ and on X, and therefore the quasi-isometry can be chosen to be truly G-equivariant.

LEMMA 3.38. *Let Δ be an ideal triangle in $\overline{\mathcal{F}}(X)$. For each $K < 0$, let Δ_K be the edges of an ideal triangle in the complete simply-connected 2-manifold of constant curvature K. For each K, let ι be the map $\iota : \Delta_K \to \Delta$, unique up to permutation of vertices, which is an isometry on each edge, and which is compatible with the nearest point parameterizations. Then for suitable K depending only on n and δ, the map ι_K is Lipschitz, with Lipschitz constant depending only on n and δ.*

PROOF. Multiplying distances by $K^{-1/2}$ scales curvature by K. On an ideal triangle in $\mathbb{H}^2$, with the nearest point parameterization, there is an estimate

$$d((a, c, t), (b, c, t)) \le e^{-t}$$

So it suffices to make K big enough so that $e^{-|K|^{-1/2}} \le \lambda$. Since λ depends only on n and δ, so does K. Since C depends only on n and δ, so does the Lipschitz constant. □

3.4.3. Spectral gap theorem. With a suitably modified definition, we can construct pleated surfaces in $\overline{\mathcal{F}}(X)$ just as we did in hyperbolic manifolds.

DEFINITION 3.39. A pleated surface (possibly with boundary) in $\overline{\mathcal{F}}(X)$ consists of the following data:

(1) a hyperbolic surface S containing a geodesic lamination L whose complementary regions are all ideal triangles, and for which $\partial S \subset L$
(2) a homomorphism $\rho : \pi_1(S) \to \mathrm{Isom}(\overline{\mathcal{F}}(X))$
(3) if $\tilde{L}$ denotes the preimage of L in the universal cover $\tilde{S}$, a map $\iota : \tilde{L} \to \overline{\mathcal{F}}(X)$, equivariant with respect to the covering space action of $\pi_1(S)$ on $\tilde{L}$ and the action of $\pi_1(S)$ on $\overline{\mathcal{F}}(X)$ by ρ, which multiplies distances by a fixed constant on each edge, and is compatible with the nearest point parameterizations.

Notice that with this definition, the image of an element of $\pi_1(\partial S)$ under ρ has infinite order, and fixes two points in ∂X. Notice too that the map ι is not typically an isometry on leaves of $\tilde{L}$, but is rather an isometry after the metric on S has been scaled by some factor. The reason is so that we can insist that the map ι is *Lipschitz*, as in Lemma 3.38.

For a given $\rho : \pi_1(S) \to \mathrm{Isom}(\overline{\mathcal{F}}(X))$ it is by no means clear which laminations L on S are realized by pleated surfaces for some hyperbolic structure on S. However, if L is *proper* (i.e. every leaf accumulates only on the boundary) then the natural analogues of Lemma 3.6 and Lemma 3.7 are valid, with essentially the same proof, at least in the case where $\rho(\pi_1(S))$ does not contain any elliptic or parabolic elements. For the sake of simplicity therefore, we state our theorems below for torsion free hyperbolic groups.

LEMMA 3.40. *Suppose* $\mathrm{Isom}(X)$ *is torsion-free and cocompact, and* $\rho : \pi_1(S) \to \mathrm{Isom}(X)$ *is incompressible (i.e. injective on essential simple loops). Then there is a pleated surface in the sense of Definition 3.39 compatible with* ρ.

PROOF. We show how to choose a hyperbolic metric on S so that ι as in Definition 3.39 exists. Explicitly, choose K as in Lemma 3.38. We will construct a metric on S of constant curvature K; scaling this metric completes the proof.

Let $g \in \pi_1(S)$ be in the conjugacy class of the loop ∂S. If l is a geodesic whose ends spiral around ∂S, the ends of a lift $\tilde{l}$ are asymptotic to two fixed points of conjugates of g. Using ρ and equivariance, the images of these fixed points in ∂X are well-defined and distinct. As in Lemma 3.7, we can choose a proper full lamination L on S (i.e. one for which every complementary region is an ideal triangle, and each geodesic spirals around ∂S at both ends) for which the three points in $\partial \pi_1(S)$ associated to each ideal triangle are mapped to three distinct points in ∂X.

For each edge of $\tilde{L}$ there is a corresponding flowline of $\overline{\mathcal{F}}(X)$ we would like to map it to. If we fix an ideal triangle of constant curvature K, there is a unique map ι from its boundary to $\overline{\mathcal{F}}(X)$ which is isometric on each edge and compatible with the nearest point parameterizations at each of the three endpoints.

An edge in $\tilde{L}$ contained in two distinct triangles in $\tilde{S}$ inherits two different parameterizations; glue the corresponding ideal triangles in $\tilde{S}$ with a shear which is the difference of these two parameterizations. Then ι as defined on the two triangles is compatible on this edge. Since L is proper, the result of this gluing is connected, and determines a (scaled) hyperbolic structure on $\tilde{S}$ and a Lipschitz map $\iota : \tilde{L} \to \overline{\mathcal{F}}(X)$ which is an isometry on each edge. This construction is equivariant, and therefore the scaled hyperbolic structure on $\tilde{S}$ covers a scaled hyperbolic structure on S. □

From this fact we can deduce analogues of Theorem 3.9 and Theorem 3.11.

THEOREM 3.41 (Calegari–Fujiwara [**49**], Thm. A). *Let G be a torsion-free group which is δ-hyperbolic with respect to a symmetric generating set $|A|$. Then there is a positive constant $C(\delta, |A|) > 0$ such that for all nontrivial $a \in G$ there is an inequality* $\mathrm{scl}(a) \geq C$.

PROOF. Let $a \in G$ be given. Let X denote the Cayley graph $C_A(G)$, and construct $\mathcal{F}(X)$ and $\overline{\mathcal{F}}(X)$. Let S be a surface of genus g with one boundary component, and $\rho : \pi_1(S) \to G$ a homomorphism taking the generator of $\pi_1(\partial S)$ to a^n. By Lemma 3.40, after reducing the genus of S if necessary, we can find a pleated surface (S, L) and $\iota : \tilde{L} \to \overline{\mathcal{F}}(X)$ with notation as in Definition 3.39. Let C be such that ι is C-Lipschitz.

As in the proof of Theorem 3.9, for any $\epsilon > 0$, we can find a component σ of $\partial S \cap S_{<\epsilon}$ of length at least

$$\mathrm{length}(\sigma) \geq \frac{\mathrm{length}(\partial S)}{12g - 6} - \frac{2\pi}{3\epsilon}$$

and a rectangular strip R of thickness $\leq \epsilon$ with σ on one side. For the sake of notation, and by analogy with Theorem 3.9, we define $\mathrm{length}(\gamma) = \mathrm{length}(\partial S)/n$.

If τ denotes the translation length of $\rho(a)$ on a flowline of $\overline{\mathcal{F}}(X)$, then $\tau \leq C \cdot \text{length}(\gamma)$. Assume $\text{length}(\sigma) > \text{length}(\gamma) + 4\epsilon$ and let p be one endpoint of σ. Let $\tilde{\sigma} \subset \tilde{L}$ be a lift of σ to $\tilde{S}$ and let $\tilde{p}$ be the corresponding lift of p. Let $a, bab^{-1} \in \pi_1(S)$ be as in the proof of Theorem 3.9. Then we have

$$d^{\times}(\iota(\tilde{p}), \rho(abab^{-1})(\iota(\tilde{p}))) \leq 4C\epsilon$$

This implies that the translation length of $\rho(abab^{-1})$ on $\overline{\mathcal{F}}(X)$ is at most $4C\epsilon$, and therefore, by bullet (6) of Theorem 3.37 the translation length of $\rho(abab^{-1})$ on X is at most $4Ck\epsilon$. On the other hand, by Corollary 3.35, since G is torsion free, there is a positive lower bound C' on the translation length of any nontrivial element of G. So if we choose ϵ so that $4Ck\epsilon < C'$ we can conclude that $\rho(bab^{-1}) = \rho(a)^{-1}$; which implies $\rho(b)$ has finite order in G, contrary to the hypothesis that G is torsion free.

This contradiction implies that $\text{length}(\sigma) \leq \text{length}(\gamma) + 4\epsilon$ and therefore

$$\text{length}(\gamma) \cdot \left(\frac{n}{12g - 6} - 1 \right) \leq 4\epsilon + \frac{2\pi}{3\epsilon}$$

On the other hand, since a is nontrivial, $Ck \cdot \text{length}(\gamma) \geq \tau(a) \geq C'$ (note that additive constants in quasi-isometries disappear when comparing translation lengths). Putting this together with our earlier estimate, and rearranging gives

$$\text{scl}(a) \geq \frac{1}{12} \left(\frac{C'}{C' + Ck \cdot \left(k\epsilon + \frac{2\pi}{3\epsilon}\right)} \right)$$

Finally, all constants which appear depend only on δ and $|A|$. □

THEOREM 3.42 (Calegari–Fujiwara [**49**], Thm. B). *Let G be a torsion-free nonelementary word hyperbolic group. Let $\delta_\infty(G)$ be the first accumulation point for stable commutator length on conjugacy classes in G. Then*

$$\frac{1}{12} \leq \delta_\infty(G) \leq \frac{1}{2}$$

PROOF. With setup and notation as in Theorem 3.41 we obtain the estimate

$$\text{length}(\gamma) \cdot \left(\frac{n}{12g - 6} - 1 \right) \leq 4\epsilon + \frac{2\pi}{3\epsilon}$$

If γ is sufficiently long, this implies $n/(12g - 6)$ is arbitrarily close to 1, so $\text{scl}(a)$ cannot be much smaller than $1/12$. This establishes the lower bound.

The upper bound follows exactly as in the proof of Theorem 3.11 by finding a quasi-isometrically embedded copy of F_2, the free group of rank 2, in G (which exists because G is nonelementary). □

3.5. Counting quasimorphisms

The geometric methods we have used to this point can be pushed only so far. The construction of Mineyev's flow space and the fine properties of its metric are very delicate and involved, and there are no realistic prospects of extending them more generally (e.g. to non-proper δ-hyperbolic spaces). Instead we turn to a generalization of Brooks' counting quasimorphisms (see § 2.3.2) due to Epstein–Fujiwara [**78**] for hyperbolic groups, and Fujiwara [**82**] in general.

3.5.1. Definition and properties. Let G be a group acting simplicially on a δ-hyperbolic complex X (not assumed to be locally finite).

DEFINITION 3.43. Let σ be a finite oriented simplicial path in X, and let σ^{-1} denote the same path with the opposite orientation. A *copy* of σ is a translate $a \cdot \sigma$ where $a \in G$.

If we fix a basepoint $p \in X$, then for any $a \in G$ there is a geodesic γ from p to $a(p)$. It is no good to try to define a counting function by counting (disjoint) copies of σ in γ, since γ is in general not unique. Instead, one considers a function which is sensitive to *all possible paths* from p to $a(p)$.

DEFINITION 3.44. Let σ be a finite oriented simplicial path in X, and let $p \in X$ be a base vertex. For any oriented simplicial path γ in X, let $|\gamma|_\sigma$ denote the maximal number of disjoint copies of σ contained in γ. Given $a \in G$, define

$$c_\sigma(a) = d(p, a(p)) - \inf_\gamma(\text{length}(\gamma) - |\gamma|_\sigma)$$

where the infimum is taken over *all* oriented simplicial paths γ in X from p to $a(p)$. Define the (small) *counting quasimorphism* h_σ by the formula

$$h_\sigma(a) = c_\sigma(a) - c_{\sigma^{-1}}(a)$$

Since length and $|\cdot|_\sigma$ take integer values on simplicial paths, the infimum of $\text{length}(\gamma) - |\gamma|_\sigma$ is *achieved* on some path γ. Any path with this property is called a *realizing path* for c_σ.

One may similarly define a "big" counting function C_σ which counts all copies of σ in each path γ, and a "big" counting quasimorphism H_σ. For the moment these are just names; we will show that h_σ is a quasimorphism, and estimate its defect in terms of δ.

If p, q are any two vertices in X, one can define

$$c_\sigma([p,q]) = d(p,q) - \inf_\gamma(\text{length}(\gamma) - |\gamma|_\sigma)$$

where the infimum is taken over all paths from p to q.

LEMMA 3.45. *One has the following elementary facts:*

(1) $c_\sigma([p,q]) = c_{\sigma^{-1}}([q,p])$
(2) $|c_\sigma([p,q]) - c_\sigma([p,q'])| \le d(q,q')$
(3) *If q is on a realizing path for σ from p to r, then*

$$c_\sigma([p,r]) \ge c_\sigma([p,q]) + c_\sigma([q,r]) \ge c_\sigma([p,r]) - 1$$

PROOF. Reversing a realizing path for c_σ gives a realizing path for $c_{\sigma^{-1}}$. A realizing path from p to q can be concatenated with a path of length $d(q,q')$ to produce some path from p to q', and *vice versa*. If q is on a realizing path from p to r, then it can intersect at most one copy of σ in that path. □

In the sequel, we always assume that the length of σ is at least 2. It follows that $\text{length}(\gamma) - |\gamma|_\sigma \ge \text{length}(\gamma)/2$ for any path γ, and we obtain an *a priori* upper bound on $d(p, a(p)) - \text{length}(\gamma) + |\gamma|_\sigma$.

Realizing paths have the following universal geometric property:

LEMMA 3.46 (Fujiwara, Lemma 3.3 [**82**]). *Suppose* $\text{length}(\sigma) \ge 2$. *Any realizing path for c_σ is a $(2,4)$-quasigeodesic.*

PROOF. Let γ be a realizing path, and q, r points on γ. Let α be the subpath of γ from q to r, and let β be a geodesic with the same endpoints. Then β intersects at most two disjoint copies of σ in γ. Let γ' be obtained from γ by cutting out the subpath from q to r and replacing it with β. We have

$$|\gamma'|_\sigma \geq |\gamma|_\sigma - 2 - |\alpha|_\sigma \geq |\gamma|_\sigma - 2 - \text{length}(\alpha)/2$$

since each copy of σ in α has length at least 2 by assumption. On the other hand, since γ is a realizing path,

$$\text{length}(\gamma') - |\gamma'|_\sigma \geq \text{length}(\gamma) - |\gamma|_\sigma$$

Since $\text{length}(\gamma') - \text{length}(\gamma) = \text{length}(\beta) - \text{length}(\alpha)$, putting these estimates together gives

$$\text{length}(\beta) \geq \text{length}(\alpha)/2 - 2$$

□

REMARK 3.47. More generally, one can obtain better constants

$$K = \frac{\text{length}(\sigma)}{\text{length}(\sigma) - 1}, \ \epsilon = \frac{2 \cdot \text{length}(\sigma)}{\text{length}(\sigma) - 1}$$

which depend explicitly on the length of σ. The argument is essentially the same as that of Lemma 3.46.

By bullet (1) from Theorem 3.30 (i.e. the "Morse Lemma"), there is a constant $C(\delta)$ such that any realizing path for c_σ from p to $a(p)$ must be contained in the C-neighborhood of any geodesic between these two points. In particular, we have the following consequence:

LEMMA 3.48. *There is a constant $C(\delta)$ such that for any path σ in X of length at least 2, and for any $a \in G$, if the C-neighborhood of any geodesic from p to $a(p)$ does not contain a copy of σ, then $c_\sigma(a) = 0$.*

Finally, the defect of h_σ can be controlled *independently* of $\text{length}(\sigma)$:

LEMMA 3.49 (Fujiwara, Prop. 3.10 [**82**]). *Let σ be a path of length at least 2. Then there is a constant $C(\delta)$ such that $D(h_\sigma) \leq C$.*

PROOF. It is evident from the definitions that h_σ is antisymmetric, so it suffices to bound $|h_\sigma(a) + h_\sigma(b) + h_\sigma(b^{-1}a^{-1})|$. More generally, let p_1, p_2, p_3 be any three points in X. We will bound $|\sum_i h_\sigma([p_i, p_{i+1}])|$ where here and in the sequel, indices are taken mod 3.

Let α_i and α_i' be realizing paths for c_σ and $c_{\sigma^{-1}}$ respectively from p_i to p_{i+1}. By δ-thinness and Lemma 3.46, we can find points q_i, q_i' in each α_i, α_i' so that all 6 points are mutually within distance $N = N(\delta)$ of each other.

By definition, $|\sum_i h_\sigma([p_i, p_{i+1}])| = |\sum_i c_\sigma(\alpha_i) - c_{\sigma^{-1}}(\alpha_i')|$. By Lemma 3.45

$$c_\sigma(\alpha_i) \geq c_\sigma([p_i, q_i]) + c_\sigma([q_i, p_{i+1}]) \geq c_\sigma(\alpha_i) - 1$$

and

$$c_{\sigma^{-1}}(\alpha_i') \geq c_{\sigma^{-1}}([p_i, q_i']) + c_{\sigma^{-1}}([q_i', p_{i+1}]) \geq c_{\sigma^{-1}}(\alpha_i') - 1$$

By Lemma 3.45 again, $|c_\sigma([q_i, p_{i+1}]) - c_\sigma^{-1}([p_{i+1}, q_{i+1}'])| \leq N$ for each i, and therefore $D \leq 6 + 6N$ by the triangle inequality and the estimates above. □

3.5.2. Weak proper discontinuity. Lemma 3.49 is *not* by itself enough to deduce the existence of nontrivial quasimorphisms on a group G acting simplicially on a δ-hyperbolic complex X, as the following example shows.

EXAMPLE 3.50. Let $G = \mathrm{SL}(2, \mathbb{Z}[1/2])$. The ring $\mathbb{Z}[1/2]$ admits a discrete 2-adic valuation, with valuation ring $\mathbb{Z}$. Let $A = \mathrm{SL}(2, \mathbb{Z})$ thought of as a subgroup of G, and let B be the group of matrices of the form

$$B = \left\{ \begin{pmatrix} a & 2^{-1}b \\ 2c & d \end{pmatrix} \right\}$$

where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is in $\mathrm{SL}(2, \mathbb{Z})$. The intersection $C = A \cap B$ is the subgroup of $\mathrm{SL}(2, \mathbb{Z})$ consisting of matrices of the form $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ where c is an even integer, and the group G is abstractly isomorphic to $A *_C B$.

There is a natural simplicial action of G on the Bass–Serre tree associated to its description as an amalgamated free product. Note that C has index 3 in both A and B, and therefore the Bass–Serre tree is regular and 3-valent. The action of G on this tree is simplicial and minimal. Nevertheless, $Q(G) = 0$, as follows from a Theorem of Liehl (see Example 5.38).

In order to show that h_σ is nontrivial, one wants to use Lemma 3.48. To apply this lemma, it is only necessary to find elements g such that if l is an axis for g on X, there are no translates of l^{-1} (i.e. l with the opposite orientation) which stay almost parallel to l on a scale large compared to the translation length of g, say of size $N\tau(g)$ where $\tau(g)$ is the translation length of g. Informally, such a pair of translates are said to be "anti-aligned". If p is in the midpoint of such a pair of anti-aligned axes l and $l' = h(l)$, the element $hg^nh^{-1}g^n$ translates $g^m(p)$ a uniformly short distance, for all n, m with $|n| + |m|$ small compared to N.

This discussion motivates the following definition, introduced in the paper [**13**] by Bestvina–Fujiwara:

DEFINITION 3.51. Suppose a group G acts simplicially on a δ-hyperbolic complex X. The action of an element $g \in G$ is *weakly properly discontinuous* if for every x and every $C > 0$ there is a constant $N > 0$ such that the set of elements $f \in G$ for which

$$d_X(x, fx) \le C \text{ and } d_X(g^N x, fg^N x) \le C$$

is finite.

To see how this addresses the issue of anti-aligned axes, suppose h_i are a sequence of elements for which $h_i(l)$ is anti-aligned with l (i.e. the axes are C apart with opposite orientation) on bigger and bigger segments centered at $x \in l$, and where C is as in Lemma 3.48. Then every f_i of the form $f_i = h_i g^n h_i^{-1} g^n$ satisfies $d_X(x, f_i x) \le C$ and $d_X(g^N x, f_i g^N x) \le C$ for any fixed N, n providing i is sufficiently big. If there are only finitely many distinct f_i, then for some i and for some distinct n, m we conclude

$$h_i g^n h_i^{-1} g^n = h_i g^m h_i^{-1} g^m$$

and therefore

$$h_i g^{n-m} h_i^{-1} = g^{m-n}$$

In other words, some nontrivial power of g is conjugate to its inverse.

Conversely, suppose no nontrivial power of g is conjugate to its inverse, and suppose that the action of g on X is weakly properly discontinuous. Let σ be a fundamental domain for the action of g on an axis, and let σ^N be a fundamental domain for g^N. Then for sufficiently big n, no translate of σ^{-N} is contained in any realizing path for g^n, and therefore the homogenization of h_{σ^N} is nontrivial on g.

Remark 3.52. To make this discussion rigorous, one must replace "axis" throughout by "quasi-axis". This extension is routine and does not lead to any more substantial difficulties. See [**13**] for details.

3.5.3. Crossing number and growth in surface groups. We briefly mention a nontrivial application of counting quasimorphisms. Let S be a closed, orientable surface of genus at least 2.

Definition 3.53. If $a \in \pi_1(S)$ is primitive, the *crossing number of* a, denoted $\mathrm{cr}(a)$, is the number of self-intersections of the geodesic representative of the free homotopy class of a in S. If $b = a^n$ then define $\mathrm{cr}(a^n) = n^2\mathrm{cr}(a)$.

Actually, crossing number is a somewhat subtle notion. For precise definitions, see § 4.2.1.

Remark 3.54. Note that the function $\mathrm{cr}(\cdot)$ is characteristic (i.e. constant on orbits of $\mathrm{Aut}(G)$).

For each non-negative integer n let $S_n \subset \pi_1(S)$ denote the set consisting of elements with $\mathrm{cr}(a) \le n$. Note that S_n generates $\pi_1(S)$ for all $n \ge 0$. For each $a \in \pi_1(S)$, let $w_n(a)$ denote the word length of a in the generators S_n.

The following is the main theorem of [**44**]:

Theorem 3.55 (Calegari [**44**], Thm. A). *Let S be a closed, orientable surface of genus at least 2. Then there are constants $C_1(S), C_2(S), C_3(S)$ such that for any non-negative integers n, m and any $a \in \pi_1(S)$ with $\mathrm{cr}(a) > 0$ there is an inequality*

$$w_n(a^m) \ge \frac{C_1 m}{\sqrt{n} + C_2} - C_3$$

A rough outline of the proof is as follows. Fix a finite generating set A for $\pi_1(S)$, and consider the Cayley graph $C_A(S)$. For each a, we build a counting quasimorphism h associated to a multiple of a which has an axis in $C_A(S)$. If $b \in \pi_1(S)$ satisfies $h(b) \ne 0$, then an axis for b contains a long segment which is close to the axis of a. This implies that the geodesic representative of b has a long segment which is close to the geodesic representative of a, and therefore b has a definite number of self-intersections. More precisely, if a realizing path for b contains p copies of a fundamental domain for the axis of a, the geodesic representative of b contains at least p^2 self-intersections. In particular, one obtains an estimate $|h(b)| \lesssim O(\sqrt{n}) + O(1)$. Since the defect of h is independent of a, b, the proof follows.

3.5.4. Separation theorem. If G is δ-hyperbolic with finite generating set $|A|$, the action of G on the Cayley graph $C_A(G)$ is properly discontinuous (and therefore certainly weakly properly discontinuous). It follows that there are many nontrivial counting quasimorphisms on G. In fact, one has the following theorem, which generalizes Theorem 3.41:

Theorem 3.56 (Calegari–Fujiwara [**49**], Thm. A′). *Let G be a group which is δ-hyperbolic with respect to some symmetric generating set A. Let a be nontorsion, with no positive power conjugate to its inverse. Let $a_i \in G$ be a collection of* *elements with $T := \sup_i \tau(a_i)$ finite. Suppose that for all nonzero integers n, m and all $b \in G$ and indices i we have an inequality*

$$a_i^m \neq ba^n b^{-1}$$

Then there is a homogeneous quasimorphism $\phi \in Q(G)$ such that

(1) *$\phi(a) = 1$ and $\phi(a_i) = 0$ for all i*
(2) *The defect satisfies $D(\phi) \leq C(\delta, |A|)\left(\frac{T}{\tau(a)} + 1\right)$*

Proof. By Lemma 3.34, after replacing each a_i by a fixed power whose size depends only on δ and $|A|$, we can assume that each a_i acts as translation on some geodesic axis l_i. Similarly, let l be a geodesic axis for a. Choose some big N (to be determined), and let σ be a fundamental domain for the action of a^N on l. The quasimorphism ϕ will be a multiple of the homogenization of h_σ, normalized to satisfy $\phi(a) = 1$. We need to show that if N is chosen sufficiently large, there are no copies of σ or σ^{-1} contained in the C-neighborhood of any l_i or l^{-1}, where C is as in Lemma 3.48.

Suppose for the sake of argument that there is such a copy, and let p be the midpoint of σ. The segment σ is contained in a translate $b(l)$. The translation length of a_i on l_i is $\tau(a_i) \leq T$, and the translation length of bab^{-1} on $b(l)$ is $\tau(a)$ (the case of l^{-1} is similar and is omitted). For big N, we can assume the length of σ is large compared to $\tau(a)$ and $\tau(a_i)$. Then for each n which is small compared to N, the element $w_n := a_i ba^n b^{-1} a_i^{-1} ba^{-n} b^{-1}$ satisfies $d(p, w_n(p)) \leq 4C$. Since there are less than $|A|^{4C}$ elements in the ball of radius $4C$ about any point, eventually we must have $w_n = w_m$ for distinct n, m. But this implies

$$a_i ba^n b^{-1} a_i^{-1} ba^{-n} b^{-1} = a_i ba^m b^{-1} a_i^{-1} ba^{-m} b^{-1}$$

and therefore a_i^{-1} and $ba^{n-m}b^{-1}$ commute. Since G is hyperbolic, commuting elements have powers which are equal, contrary to the hypothesis that no conjugate of a has a power equal to a power of a_i.

This contradiction implies that $\tau(a_i) + |A|^{4C}\tau(a) \geq N\tau(a)$. On the other hand, $D(h_\sigma)$ is uniformly bounded, by Lemma 3.49, and satisfies $h_\sigma(a^{Nn}) \geq n$. Homogenizing and scaling by the appropriate factor, we obtain the desired result. □

In fact, let $\sum n_i a_i$ be any integral chain which is nonzero in B_1^H. Without loss of generality, we may replace this chain by a rational chain with bounded denominators, with the same scl, and such that no distinct a_i, a_j have conjugate powers. After reordering, suppose $\tau(a_1) \geq \tau(a_i)$ for all i, and let ϕ be as in Theorem 3.56, so that $\phi(a_1) = 1$ and $\phi(a_i) = 0$ for $i \neq 1$. The defect $D(\phi)$ is bounded above by a constant depending only on δ and $|A|$. The coefficient of a_1 is bounded below by a positive constant depending only on δ and $|A|$. Hence by Bavard duality, $\text{scl}(\sum n_i a_i)$ is bounded below by a positive constant depending only on δ and $|A|$. In other words we have proved:

Corollary 3.57. *Let G be hyperbolic. Then* scl *is a norm on $B_1^H(G)$. Moreover, the value of* scl *on any nonzero integral chain in $B_1^H(G)$ is bounded below by a positive constant that depends only on δ and $|A|$.*

3.6. Mapping class groups

In this section we survey some of what is known about scl in mapping class groups. Our survey is very incomplete, since our main goal is to state an analogue of Theorem 3.41 for mapping class groups, and to give the idea of the proof.

DEFINITION 3.58. Let S be an oriented surface (possibly punctured). The *mapping class group* of S, denoted $\text{MCG}(S)$, is the group of isotopy classes of orientation-preserving self-homeomorphisms of S.

EXAMPLE 3.59 (Dehn twist). Let γ be an essential simple curve in S. A *right-handed Dehn twist* in γ is the map $t_\gamma : S \to S$ supported on an annulus neighborhood $\gamma \times [0,1]$ which takes each curve $\gamma \times t$ to itself by a positive twist through a fraction t of its length. If the annulus is parameterized as $\mathbb{R}/\mathbb{Z} \times [0,1]$, then in co-ordinates, the map is given by $(\theta, t) \to (\theta + t, t)$.

By abuse of notation, we typically refer to both a specific homeomorphism and its image in $\text{MCG}(S)$ as a Dehn twist.

REMARK 3.60. The inverse of a right-handed Dehn twist is a left-handed Dehn twist. Sometimes, right-handed Dehn twists are called *positive* Dehn twists. Notice that the handedness of a Dehn twist depends on an orientation for S but *not* on an orientation for γ.

The mapping class group is a fundamental object in 2-dimensional topology, and the literature on it is vast. Our treatment of it in this section is very cursory, and intended mainly just to introduce definitions and notation. For simplicity, we restrict attention throughout this section to mapping class groups of closed, orientable surfaces, although most of the results generalize to mapping class groups of surfaces with boundary or punctured surfaces. We refer the interested reader to [**16**] or [**80**] for background and details.

An element of $\text{MCG}(S)$ induces an outer automorphism of $\pi_1(S)$ (outer, because homeomorphisms are not required to keep the basepoint fixed). This fact connects geometry with algebra. In fact, the connection is more intimate than it may appear at first glance, because of

THEOREM 3.61 (Dehn–Nielsen). *The natural map* $\text{MCG}(S) \to \text{Out}(\pi_1(S))$ *is an injection, with image equal to subgroup consisting of automorphisms which permute the peripheral subgroups.*

In particular, for S closed, $\text{MCG}(S)$ is isomorphic to $\text{Out}(\pi_1(S))$. For each S, the group $\text{MCG}(S)$ is finitely presented. A finite generating set consisting of Dehn twists was first given by Dehn; a description of a finite set of relations (with respect to a slightly different generating set) was first given by Hatcher–Thurston [**106**].

The mapping class group of any closed orientable S is generated by finite order elements; in particular, its Abelianization is finite. If the genus of S is at least 3, the Abelianization is trivial:

THEOREM 3.62 (Powell, [**171**]). *Let S be a closed, orientable surface of genus at least* 3. *Then* $\text{MCG}(S)$ *is perfect.*

A short proof of this theorem is due to Harer:

PROOF. It is well-known that $\text{MCG}(S)$ is generated by Dehn twists about nonseparating curves (e.g. Lickorish's generating set [**134**]). By the classification of surfaces, any two nonseparating curves may be interchanged by a homeomorphism of S; it follows that $H_1(\text{MCG}(S);\mathbb{Z})$ is generated by the image t of a twist about any nonseparating curve.

If the genus of S is at least 3, then S contains a non-separating four-holed sphere. The *lantern relation*, in MCG(4-holed sphere), says that the product of Dehn twists in the boundary components of a 4-holed sphere is equal to the product of twists in three curves in the sphere which separate the boundary components in pairs, and intersect each other in two points. The image of this relation in $H_1(\text{MCG}(S);\mathbb{Z})$ is $t^4 = t^3$, so $\text{MCG}(S)$ is perfect. □

3.6.1. Right-handed Dehn twists. Interesting lower bounds in scl can be obtained using gauge theory. This is a subject which has been pioneered by Kotschick, in [**130, 131**] and Endo–Kotschick [**73, 74**].

The following theorem is essentially due to Endo–Kotschick [**73**] although for technical reasons, the result is stated in that paper only for powers of a single separating Dehn twist. This technical assumption is removed in [**23**], and the result extended to products of positive twists in disjoint simple curves in [**131**].

THEOREM 3.63 (Endo–Kotschick [**73**], Kotschick [**131**]). *Let S be a closed orientable surface of genus $g \geq 2$. If $a \in \text{MCG}(S)$ is the product of k right-handed Dehn twists along essential disjoint simple closed curves $\gamma_1, \cdots, \gamma_k$ then*

$$\text{scl}(a) \geq \frac{k}{6(3g-1)}$$

It is beyond the scope of this survey to give a complete proof, but the way in which gauge theory enters the picture is the following. The product $a = t_{\gamma_1} t_{\gamma_2} \cdots t_{\gamma_k}$ lets one build a Lefschetz fibration E over the disk with fiber S which is singular over k distinct points, and such that the restriction of E to ∂D is a surface bundle with monodromy a. Over each singular point p_i, the fiber is a copy of S "pinched" along the curve γ_i, and such that the monodromy of a small loop around p_i is the twist t_{γ_i}. Since the curves γ_i are all disjoint, we can adjust the fiber structure on E so there is only *one* singular fiber, which degenerates along all the γ_i simultaneously. Since the twists are all right-handed, E admits a symplectic structure. Take an n-fold branched cover of the disk over the singular point, and pull back the fibration. After a suitable resolution, we get a new symplectic Lefschetz fibration E' over D with one singular fiber, such that the monodromy around the boundary is a^n, and such that the singular fiber has kn vanishing cycles, which come in parallel families of the γ_i.

An expression of a^n as a product of commutators in $\text{MCG}(S)$ defines a nonsingular S bundle E'' over a once-punctured surface F, and by gluing E'' to E' along their boundaries in a fiber-respecting way, one obtains a closed symplectic manifold W. Then the engine of the proof is the well-known theorem of Taubes [**194**] in Seiberg–Witten theory which shows that for a minimal symplectic 4-manifold with $b_2^+ > 1$ the canonical class is represented by a symplectically embedded surface without spherical components. From this one derives inequalities on intersection numbers of certain surfaces in W and the result follows.

It is crucial in Theorem 3.63 that the twists in the different curves should all have the same handedness.

EXAMPLE 3.64 (Kotschick, Endo–Kotschick [**131, 74**]). Let α be an essential simple closed curve, and let $g \in \mathrm{MCG}(S)$ be such that $g(\alpha) \cap \alpha = \emptyset$, and $g(\alpha)$ is not isotopic to α. Let $h = t_\alpha t_{g(\alpha)}^{-1}$. Since α and $g(\alpha)$ are disjoint,

$$h^n = t_\alpha^n t_{g(\alpha)}^{-n} = t_\alpha^n g t_\alpha^{-n} g^{-1} = [t_\alpha^n, g]$$

so $\mathrm{scl}(h) = 0$. Note in this case that there is always some $f \in \mathrm{MCG}(S)$ which interchanges α and $g(\alpha)$. For such an f we have $fhf^{-1} = h^{-1}$, so that $h = 0$ in B_1^H.

As another example, let α, β, γ be disjoint nonseparating non-isotopic simple closed curves, and define $h = t_\alpha^{-1} t_\beta^{-1} t_\gamma^2$. If g interchanges α and γ, and g' interchanges β and γ, then $h^k = [t_\gamma^k, g][t_\gamma^k, g']$. In this case, all powers of h are in distinct conjugacy classes, and h is not in B_1^H. This example shows that H is not closed in $B_1(\mathrm{MCG}(S))$.

Interesting upper bounds on scl can be obtained by explicit examples.

EXAMPLE 3.65 (Korkmaz [**129**]). Let $a \in \mathrm{MCG}(S)$ be a Dehn twist in a nonseparating closed curve. Then a^{10} can be written as a product of two commutators.

Let $a_1, \cdots, a_5$ be curves on S as in Figure 3.4, where a_4, a_5 are nonseparating. For each i, let t_i denote a positive Dehn twist in a_i. Notice that t_1, t_3, t_4, t_5 all

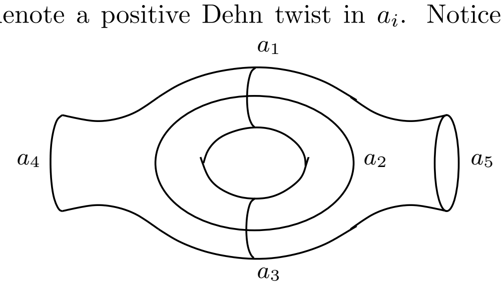


FIGURE 3.4. The curves $a_1, \cdots, a_5$ in S

commute. Moreover, a neighborhood of $a_1 \cup a_2$ is a once-punctured torus. An element of the mapping class group of a punctured torus is determined by its action on homology, and one may verify that the relation $t_1 t_2 t_1 = t_2 t_1 t_2$ holds by an elementary calculation, and similarly with t_3 in place of t_1. The relation $t_4 t_5 = (t_1 t_2 t_3)^4$ is a little harder to see, but still elementary.

Following [**129**], we calculate

$$\begin{aligned} t_4 t_5 &= (t_1 t_2 t_3)(t_1 t_2 t_3)(t_1 t_2 t_3)(t_1 t_2 t_3) \\ &= (t_1 t_2 t_1)(t_3 t_2 t_3)(t_1 t_2 t_1)(t_3 t_2 t_3) \\ &= (t_2 t_1 t_2)(t_2 t_3 t_2)(t_2 t_1 t_2)(t_2 t_3 t_2) \end{aligned}$$

Since t_2 commutes with both t_4 and t_5, this gives

$$t_4 t_5 = t_1 (t_2^2 t_3 t_2^{-2}) t_2^4 t_1 t_2^{-1} (t_2^3 t_3 t_2^{-3}) t_2^{-1} t_2^6$$

If $\alpha = t_2^2(a_3)$ and $\beta = t_2^3(a_3)$ then this yields

$$(t_4 t_\alpha^{-1} t_5 t_1^{-1}) = t_2^4 (t_1 t_2^{-1} t_\beta t_2^{-1}) t_2^6$$

Each bracketed expression is of the form $t_a t_b^{-1} t_c t_d^{-1}$ for simple curves a, b, c, d. It can be verified in each case that the curves $a \cup b$ lie in S in the same combinatorial pattern as $d \cup c$. Therefore there is $g \in \mathrm{MCG}(S)$ for which $g(a) = d$ and $g(b) = c$. But this means

$$t_a t_b^{-1} t_c t_d^{-1} = t_a t_b^{-1} g t_b t_a^{-1} g^{-1} = [t_a t_b^{-1}, g]$$

That is, each of the two bracketed expressions are commutators, and the proof follows.

In case genus$(S) = 2$, one obtains

$$\frac{1}{30} \le \mathrm{scl}(t) \le \frac{3}{20}$$

where the first inequality comes from Theorem 3.63.

Let t denote a Dehn twist in a nonseparating curve (for concreteness). One may ask to what extent $\mathrm{scl}(t)$ depends on genus(S). In fact, it turns out that Theorem 3.63 gives the correct order of magnitude. This follows from a general phenomenon, especially endemic in transformation groups, which we describe.

Example 3.66 (Münchhausen trick). Suppose we are given a group G acting on a set Y. Suppose further that there is an identity $a = \prod_{i=1}^n [b_i, c_i]$ in G where a, b_i, c_i all have support in some subset $X \subset Y$. Suppose finally that there is $g \in G$ such that $X \cap g^i(X) = \emptyset$ for $0 < i \le m$.

If H is a subgroup of G consisting of elements with support in X, define $\Delta : H \to G$ by

$$\Delta(h) = h h^g h^{g^2} \cdots h^{g^{m-1}}$$

where the superscript notation denotes conjugation. The condition on g ensures Δ is a homomorphism, and therefore $\mathrm{cl}(\Delta(a)) \le n$. Now define an element $j = a(a^2)^g (a^3)^{g^2} \cdots (a^m)^{g^{m-1}}$. We have the identity $[j, g] = \Delta(a)(a^{-m})^{g^m}$, which exhibits a^m as a product of at most $n+1$ commutators. If m is large compared to n, then $\mathrm{scl}(a)$ is small.

Corollary 3.67 (Kotschick). *If t is a Dehn twist in a non-separating curve, there is an estimate* $\mathrm{scl}(t) = O(1/g)$.

Proof. The lower bound is Theorem 3.63. The upper bound follows by exhibiting t as a product of commutators of elements b_i, c_i supported in some fixed surface T with boundary, which can be included into a surface S of arbitrary genus. Then apply the Münchhausen trick. □

Remark 3.68. Many variations on Corollary 3.67 are proved in [**132**] and [**29**], and the same trick appears in the proof of Theorem 5.13. The trick works whenever there is "enough room" in Y for many disjoint copies of X; in many important applications, X is (in some sense) a copy of Y. The terminology "Münchhausen trick" is taken from [**118**], and "refers to the story about how the legendary baron allegedly succeeded in pulling himself out of a quagmire by his own hair". This trick goes back at least to [**81**] (in fact one could argue it goes back to Zeno of Elea).

3.6.2. The complex of curves. For most surfaces S, the group $\mathrm{MCG}(S)$ is not word-hyperbolic. Nevertheless, it acts naturally on a certain δ-hyperbolic simplicial complex, called the *complex of curves*. This complex was first introduced by Harvey [**103**], but it was Masur–Minsky [**148**] who established some of its most important basic properties. A similar complex was also introduced by Hatcher–Thurston [**106**]. A good introductory reference to the complex of curves is [**182**].

DEFINITION 3.69 (Harvey [**103**]). Let S be a closed, orientable surface of genus at least 2. The *complex of curves*, denoted $\mathcal{C}(S)$, is the simplicial complex whose k simplices consist of isotopy classes of pairwise disjoint non-parallel essential simple closed curves on S.

With this definition, $\mathcal{C}(S)$ is a simplicial complex of dimension $3g-4$. The natural permutation action of $\mathrm{MCG}(S)$ on the set of isotopy classes of essential simple curves on S induces a simplicial action of $\mathrm{MCG}(S)$ on $\mathcal{C}(S)$.

REMARK 3.70. Similar definitions can be made when S has smaller genus, or has punctures or boundary components. See [**148**].

We can think of $\mathcal{C}(S)$ as a metric space, by taking every edge to have length 1 and every simplex to be equilateral. In the sequel, we are typically interested not in $\mathcal{C}(S)$ itself, but in its 1-skeleton. Usually, by abuse of notation, when we talk about $\mathcal{C}(S)$ we really mean its 1-skeleton. It should be clear from context which sense is meant in each case.

The main property of $\mathcal{C}(S)$ from our point of view is the following theorem:

THEOREM 3.71 (Masur–Minsky [**148**]). *Let S be as above. Then $\mathcal{C}(S)$ is δ-hyperbolic for some $\delta(S)$.*

An element $a \in \mathrm{MCG}(S)$ is *reducible* if it permutes some finite set of isotopy classes of disjoint, non-parallel essential simple closed curves. It turns out that an element $a \in \mathrm{MCG}(S)$ has a finite orbit in $\mathcal{C}(S)$ if and only if a is either finite order or reducible. An element which is neither finite order nor reducible is said to be *pseudo-Anosov*.

THEOREM 3.72 (Masur–Minsky [**148**]). *Let $a \in \mathrm{MCG}(S)$ be pseudo-Anosov. Then every orbit of a on $\mathcal{C}(S)$ is a quasigeodesic.*

In particular, every pseudo-Anosov element has a positive translation length $\tau(a)$. In fact, Bowditch [**21**] proves the following analogue of Lemma 3.34:

THEOREM 3.73 (Bowditch, Theorem 1.4 [**21**]). *Let S be a closed, orientable surface of genus at least 2. Then there is a constant $C(S)$ such that for every pseudo-Anosov $a \in \mathrm{MCG}(S)$, there is $n \le C$ such that a^n fixes some bi-infinite geodesic axis l_a and acts on it by translation.*

3.6.3. Acylindricity. The action of $\mathrm{MCG}(S)$ on $\mathcal{C}(S)$ is not proper; the stabilizer of a vertex is isomorphic to a copy of $\mathrm{MCG}(S')$ for some smaller surface S'. Nevertheless, Bestvina–Fujiwara ([**13**]) show that every pseudo-Anosov element of $\mathrm{MCG}(S)$ acts weakly properly discontinuously on $\mathcal{C}(S)$. As a corollary, they deduce the following theorem:

THEOREM 3.74 (Bestvina–Fujiwara, [**13**], Theorem 12). *Let G be a subgroup of $\mathrm{MCG}(S)$ which is not virtually Abelian. Then the dimension of $Q(G)$ is infinite.*

In particular, if ϕ is pseudo-Anosov, and p, q are sufficiently far apart on an axis for ϕ, only finitely many elements of $\mathrm{MCG}(S)$ move both p and q a bounded distance. This is enough to show that every pseudo-Anosov element either has a (bounded) power conjugate to its inverse, or has positive scl. To obtain *uniform* estimates on scl, one needs a slightly stronger statement, captured in the following theorem of Bowditch:

THEOREM 3.75 (Bowditch, Acylindricity Theorem [**21**]). *Let S be a closed orientable surface of genus $g \ge 2$. For any $t > 0$ there exist positive constants $C_1(t,S), C_2(t,S)$ such that given any two points $x, y \in \mathcal{C}(S)$ with $d(x,y) \ge C_1$ there are at most C_2 elements $a \in \mathrm{MCG}(S)$ such that $d(x,ax) \le t$ and $d(y,ay) \le t$.*

REMARK 3.76. A similar theorem is also proved by Masur–Minsky [**148**].

We are now in a position to state the analogue of Theorem 3.41 and Theorem 3.56 for mapping class groups.

THEOREM 3.77 (Calegari–Fujiwara [**49**], Thm. C). *Let S be a closed orientable surface of genus at least 2. Then there are constants $C_1(S), C_2(S) > 0$ such that for any pseudo-Anosov element $a \in \mathrm{MCG}(S)$ either there is a positive integer $n \le C_1$ for which a^n is conjugate to its inverse, or else there is a homogeneous quasimorphism $\phi \in Q(\mathrm{MCG}(S))$ with $\phi(a) = 1$ and $D(\phi) \le C_2$.*

Moreover, suppose $a_i \in \mathrm{MCG}(S)$ are a (possibly infinite) collection of elements with $T := \sup_i \tau(a_i)$ finite. Suppose that for all nonzero integers n, m and all $b \in \mathrm{MCG}(S)$ and indices i we have an inequality

$$a_i^m \neq ba^nb^{-1}$$

Then there is a homogeneous quasimorphism $\phi \in Q(\mathrm{MCG}(S))$ such that

(1) *$\phi(a) = 1$ and $\phi(a_i) = 0$ for all i*
(2) *The defect satisfies $D(\phi) \le C_2(S)\left(\frac{T}{\tau(a)} + 1\right)$*

PROOF. The proof is essentially the same as the proof of Theorem 3.56, with Theorem 3.73 used in place of Lemma 3.34. After replacing a and a_i by (bounded) powers, one assumes that they stabilize axes l and l_i respectively. For each n, let $w_n := a_iba^nb^{-1}a_i^{-1}ba^{-n}b^{-1}$. If $b(l)$ is close to l_i on a segment σ which is long compared to $\tau(a), \tau(a_i)$ and C_1 (as in Theorem 3.75), then one can find points p and p' on l_i with $d(p,p') \ge C_1$ such that $d(p, w_n(p)) \le t$ and $d(p', w_n(p')) \le t$ for all n small compared to length(σ).

One needs to know that two pseudo-Anosov elements in $\mathrm{MCG}(S)$ which commute have powers which are proportional (the pseudo-Anosov hypothesis cannot be omitted here); see e.g [**198**]. Otherwise, the remainder of the proof is copied verbatim from the proof of Theorem 3.56. □

REMARK 3.78. In contrast with the case of word-hyperbolic groups, it should be noted that there are infinitely many conjugacy classes of pseudo-Anosov elements in $\mathrm{MCG}(S)$ with bounded translation length. In fact, the first accumulation point for translation length in $\mathrm{MCG}(S)$ is $O(1/g\log(g))$, where g is the genus of S; see Theorem 1.5 of [**79**].

The separation property of the quasimorphisms produced by Theorem 3.77 is very powerful, and has a number of consequences, including the following.

COROLLARY 3.79. *Let Σ be a subset of $\mathrm{MCG}(S)$ consisting only of reducible elements, and let G be the subgroup it generates. Suppose G contains a pseudo-Anosov element a with no power conjugate to its inverse. Then the Cayley graph of G with respect to the generating set Σ has infinite diameter.*

PROOF. By Theorem 3.77 there is a homogeneous quasimorphism ϕ defined on $\mathrm{MCG}(S)$ with $\phi(a^n) = n$ which vanishes on Σ. If b is an element of G with length at most m in the generators Σ, then $\phi(b) \le (m-1)D(\phi)$. □

The hypotheses of this Corollary are satisfied whenever G is not reducible or virtually cyclic.

Example 3.80 (Broaddus–Farb–Putman [**25**]). The *Torelli group*, denoted $\mathcal{I}(S)$, is the kernel of the natural map $\mathrm{MCG}(S) \to \mathrm{Aut}(H_1(S)) = \mathrm{Sp}(2g, \mathbb{Z})$. It is not a perfect group; the kernel of the map

$$\mathcal{I}(S) \to H_1(\mathcal{I}(S); \mathbb{Z})/\text{torsion}$$

is denoted $\mathcal{K}(S)$, and is generated by Dehn twists about separating simple closed curves. This is an infinite (in fact, characteristic) generating set. On the other hand, by Corollary 3.79 the diameter of the Cayley graph of $\mathcal{K}$ with respect to this generating set is infinite.

3.7. Out(F_n)

Very recently, the methods discussed in this chapter have been used to construct many nontrivial quasimorphisms on $\mathrm{Out}(F_n)$, the group of outer automorphisms of a free group. The main results described in § 3.7.2 were announced by Hamenstädt in May 2008, and first appeared in (pre-)print in Bestvina–Feighn [**12**]. In what follows we restrict ourselves to describing the construction of suitable δ-hyperbolic simplicial complexes on which $\mathrm{Out}(F_n)$ acts, and summarizing the important properties of these complexes and the action without justification.

3.7.1. Outer space.

In what follows, $\mathrm{Out}(F_n)$ denotes the outer automorphism group of the free group F_n of rank $n \geq 2$. The modern theory of $\mathrm{Out}(F_n)$ is dominated by several deep analogies between this group and mapping class groups. The cornerstone of these analogies is Culler–Vogtmann's construction [**61**] of *Outer space*, which serves as an analogue of Teichmüller space.

Definition 3.81. Fix F_n, a free group of rank n. An action of F_n on an $\mathbb{R}$-tree T is *minimal* if there is no proper F_n-invariant subtree of T. Let $\rho : F_n \to \mathrm{Isom}(T)$ be an action which is minimal, free and discrete. Associated to any such ρ there is a *length function* $\ell_\rho \in \mathbb{R}^{F_n}$ where $\ell_\rho(g)$ is the translation length of $\rho(g)$ on T.

Outer space, denoted in the sequel $\mathcal{PT}$, is the projectivization of the space of length functions of minimal, free, discrete actions of F_n on $\mathbb{R}$-trees, with the weak topology. Its compactification $\overline{\mathcal{PT}}$ is obtained by adding weak limits of projective classes of length functions.

If $\rho : F_n \to \mathrm{Isom}(T)$ is a minimal, free, discrete action of F_n on an $\mathbb{R}$-tree, and $\varphi : F_n \to F_n$ is an automorphism, then $\rho \circ \varphi^{-1} : F_n \to \mathrm{Isom}(T)$ is another action. If φ is inner, the length functions ℓ_ρ and $\ell_{\rho\circ\varphi^{-1}}$ are equal. Hence the group $\mathrm{Out}(F_n)$ acts in a natural way on $\mathcal{PT}$, and this action extends continuously to its compactification.

Outer space has a natural cellular structure, which can be described as follows. For each action $\rho : F_n \to \mathrm{Isom}(T)$, let Γ_ρ be the quotient of T by $\rho(F_n)$, thought of as a metric graph together with an isomorphism of its fundamental group with F_n (i.e. a *marking*), which is well-defined up to conjugacy. The cells of $\mathcal{PT}$ are the actions which correspond to a fixed combinatorial type of Γ_ρ, together with a choice of marking. This cellular structure extends naturally to $\overline{\mathcal{PT}}$.

3.7.2. Fully irreducible automorphisms.

DEFINITION 3.82. An element $\varphi \in \mathrm{Out}(F_n)$ is *fully irreducible* if for all proper free factors F of F_n and all $k > 0$ the subgroup $\varphi^k(F)$ is not conjugate to F.

The main result of [**12**] is as follows:

THEOREM 3.83 (Bestvina–Feighn [**12**], p.11). *For any finite set $\varphi_1, \cdots, \varphi_k$ of fully irreducible elements of $\mathrm{Out}(F_n)$ there is a connected δ-hyperbolic graph X (depending on the φ_i) together with an isometric action of $\mathrm{Out}(F_n)$ on X such that*

1. *the stabilizer in $\mathrm{Out}(F_n)$ of a simplicial tree in $\overline{\mathcal{PT}}$ has bounded orbits*
2. *the stabilizer in $\mathrm{Out}(F_n)$ of a proper free factor $F \subset F_n$ has bounded orbits*
3. *the φ_i all have nonzero translation lengths*

The construction of the graph X is somewhat complicated, and follows a template developed by Bowditch [**20**] to study convergence group actions. A fully irreducible automorphism ψ has one stable and one unstable fixed point in the boundary of $\overline{\mathcal{PT}}$, which we denote $T_\psi^\pm$. A tree T is *irreducible* if it is of the form T_ψ^+ for some fully irreducible ψ.

Choose sufficiently small closed neighborhoods $D_i^\pm$ of $T_{\varphi_i}^\pm$. In $\overline{\mathcal{PT}}$, let $\mathcal{M}$ be the subspace of *all* irreducible trees. Define an *annulus* to be an ordered pair of closed subsets of $\mathcal{M}$ either of the form $(\psi(D_i^-)\cap\mathcal{M}, \psi(D_i^+)\cap\mathcal{M})$ or $(\psi(D_i^+)\cap\mathcal{M}, \psi(D_i^-)\cap\mathcal{M})$, where $\psi \in \mathrm{Out}(F_n)$ and $D_i^\pm$ are as above. Denote the set of annuli (defined as above) by $\mathcal{A}$. The pair $(\mathcal{M}, \mathcal{A})$ depends on the choice of the φ_i, and both $\mathcal{M}$ and $\mathcal{A}$ admit natural actions by $\mathrm{Out}(F_n)$.

For any subset $K \subset \mathcal{M}$ and any annulus $A = (A^-, A^+)$ write $K < A$ if $K \subset \mathrm{int}A^-$, and write $A < K$ if $K \subset \mathrm{int}A^+$. If $A = (A^-, A^+)$ and $B = (B^-, B^+)$ are two annuli, write $A < B$ if $\mathrm{int}A^+ \cup \mathrm{int}B^- = \mathcal{M}$. Then for any pair of subsets K, L of $\mathcal{M}$, define $(K|L) \in [0, \infty]$ to be the biggest number of annuli A_i in $\mathcal{A}$ such that

$$K < A_1 < A_2 < \cdots < A_n < L$$

Let $\mathcal{Q}$ denote the set of ordered triples of distinct points in $\mathcal{M}$. If $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$ are elements of $\mathcal{Q}$, then define

$$\rho(A, B) = \max(\{a_i, a_j\}|\{b_k, b_l\})$$

where the maximum is taken over all $i \neq j$ and $k \neq l$. The *Bowditch complex* of the pair $(\mathcal{M}, \mathcal{A})$, is the graph whose vertices are the elements of $\mathcal{Q}$ and whose edges are the pairs of elements $A, B \in \mathcal{Q}$ with $\rho(A, B) \leq r$ for some sufficiently big r. Bowditch [**20**] gives certain axioms for an abstract pair $(\mathcal{M}, \mathcal{A})$ which ensure that the associated Bowditch complex is δ-hyperbolic. The substance of Theorem 3.83 is the proof that $(\mathcal{M}, \mathcal{A})$ as above satisfies Bowditch's axioms.

In order to construct quasimorphisms, one must also know that many elements of $\mathrm{Out}(F_n)$ act weakly properly discontinuously. This is the following proposition, also from [**12**]:

PROPOSITION 3.84 (Bestvina–Feighn [**12**], p.24). *For φ_i and X as in Theorem 3.83, the action of each φ_i on X is weakly properly discontinuous; i.e. for every $x \in X$ and every $C > 0$ there is a constant $N > 0$ such that the set of $\psi \in \mathrm{Out}(F_n)$ for which*

$$d_X(x, \psi x) \leq C \text{ and } d_X(\varphi_i^N x, \psi\varphi_i^N x) \leq C$$

is finite.

Theorem 3.83 allows one to construct many quasimorphisms on $\mathrm{Out}(F_n)$ by the method of § 3.5. Proposition 3.84 implies that these quasimorphisms are nontrivial and independent. Consequently, one concludes that $Q(\mathrm{Out}(F_n))$ is infinite dimensional; in fact ([**12**] Corollary 4.28), for any subgroup Γ of $\mathrm{Out}(F_n)$ which contains two independent fully irreducible automorphisms, $Q(\Gamma)$ is infinite dimensional.

CHAPTER 4

Free and surface groups

In this chapter we study scl in free groups, and some related groups. The methods are largely geometric and depend on realizing the groups in question as fundamental groups of particularly simple low-dimensional manifolds.

The first main theorem proved in this chapter is the Rationality Theorem (Theorem 4.24), which says that in a free group F, the unit ball of the scl norm on $B_1^H(F)$ is a rational polyhedron; i.e. scl is a piecewise linear rational function on finite dimensional rational subspaces of $B_1^H(F)$. It follows that scl takes on only rational values in free groups. The method of proof is direct: we show how to explicitly construct extremal surfaces bounding finite linear combinations of conjugacy classes. As a byproduct, we obtain a polynomial-time algorithm to calculate scl in free groups, which can be practically implemented, at least in some simple cases. This algorithm gives an interesting conjectural picture of the spectrum of scl on free groups, and perhaps some insight into the spectrum of scl on word-hyperbolic groups in general.

The polyhedrality of the unit ball of the scl norm is related to certain rigidity phenomena. Each nonzero element in $B_1^H(F)$ projectively intersects the boundary of the unit ball of the scl norm in the interior of some face. The smaller the codimension of this face, the smaller the space of quasimorphisms which are extremal for the given element. The situations displaying the most rigidity are therefore associated to faces of the unit ball of codimension one. It turns out that for a free group, such faces of codimension one exist, and have a geometric meaning. In § 4.2 we discuss the Rigidity Theorem (Theorem 4.78), which says that if F is a free group, associated to each isomorphism $F \to \pi_1(S)$ (up to conjugacy), where S is a compact oriented surface, there is a top dimensional face π_S of the unit ball of the scl norm on F, and the unique homogeneous quasimorphism ϕ_S dual to π_S (up to scale and elements of H^1) is the rotation quasimorphism associated to a hyperbolic structure on S.

Finally, in § 4.3, we discuss diagrammatic methods to study scl in free groups. In particular, we discuss a technique due to Duncan–Howie which uses left-invariant orders on one-relator groups to obtain sharp lower bounds on scl in free groups.

Some of the material in this chapter is developed more fully in the papers [**47, 43, 45, 46**].

4.1. The Rationality Theorem

The goal of this section is to prove the Rationality Theorem for free groups. Essentially, this theorem says that the unit ball in the scl norm on $B_1^H(F)$ is a rational polyhedron. Polyhedral norms occur in other contexts in low-dimensional topology, and the best-known example is that of the Thurston norm on the 2-dimensional homology of a 3-manifold. We briefly discuss this example.

4.1.1. Thurston norm. Let M be a 3-manifold. Thurston [**196**] defined a pseudo-norm on $H_2(M, \partial M; \mathbb{R})$ as follows.

For each properly embedded surface S in M, define $\|S\|_T = -\chi^-(S)$. For each relative class $A \in H_2(M, \partial M; \mathbb{Z})$, define

$$\|A\|_T = \inf_S -\chi^-(S)$$

where the infimum is taken over all properly embedded surfaces S for which $[S]$ represents the class A. Thurston shows that this function satisfies the following two crucial properties:

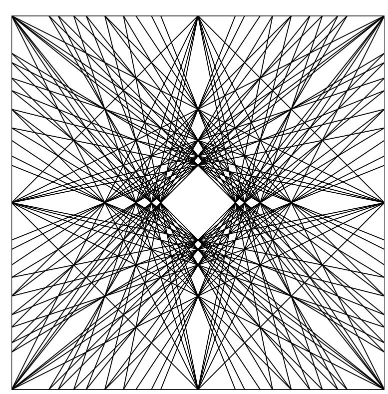
- it is linear on rays; that is, $\|nA\|_T = n\|A\|_T$ for any integral class A and any non-negative integer n
- it is subadditive; that is, $\|A+B\|_T \leq \|A\|_T + \|B\|_T$ for all integral classes A, B.

By the first property, $\|\cdot\|_T$ can be extended by linearity to all of $H_2(M, \partial M; \mathbb{Q})$. By the second property, it can be extended to a unique continuous function on $H_2(M, \partial M; \mathbb{R})$, which is linear on rays and subadditive. Such a function satisfies the axioms of a (pseudo)-norm, and is called the *Thurston norm* on homology. Note that this function is generally only a pseudo-norm; it takes the value 0 on the span of integral classes which can be represented by surfaces of non-negative Euler characteristic. If M is irreducible and atoroidal, $\|\cdot\|_T$ is a genuine norm.

By construction, $\|A\|_T \in \mathbb{Z}$ for all $A \in H_2(M, \partial M; \mathbb{Z})$. A norm on a finite dimensional vector space which takes integer values on integer vectors (with respect to some basis) can be characterized in a finite amount of data, as follows.

LEMMA 4.1. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ which takes integer values on the lattice $\mathbb{Z}^n$. Then the unit ball of $\|\cdot\|$ is a finite sided polyhedron whose faces are defined by integral linear equalities.*

PROOF. Let U be any open set in $\mathbb{R}^n$ containing 0. We claim that there are only *finitely many* integral linear functions ϕ on $\mathbb{R}^n$ such that the subspace $\phi \leq 1$ contains U. Let ϕ be such a linear function. Then there is a (unique) *integral* vector v_ϕ such that $\phi(w) = \langle v_\phi, w\rangle$ where $\langle\cdot,\cdot\rangle$ denotes the ordinary inner product on $\mathbb{R}^n$. Since U is open, there is some positive number ϵ such that the ball of radius ϵ in the (ordinary) L^1 norm is contained inside U. Hence if ϕ is as above, every co-ordinate of v_ϕ has absolute value at most $1/\epsilon$. On the other hand, since v_ϕ is integral, there are only *finitely many* functions ϕ with this property (the adjacent Figure shows all level sets $\phi = 1$ in 2 dimensions for $\epsilon = 1/5$). This proves the claim.



Let B denote the unit ball in the $\|\cdot\|$ norm. For the remainder of the proof we assume $n = 3$ (the general case is not significantly more complicated). For each integral basis $\{v_1, v_2, v_3\}$, there is a unique integral linear function on $\mathbb{R}^3$ that agrees with $\|\cdot\|$ on the elements of the basis. Pick some primitive integral vector v_1, and then extend v_1 to an integral basis $\{v_1, v_2, v_3\}$. For each pair of integers i, j let $v_2^i = v_2 + iv_1$ and $v_3^{i,j} = v_3 + iv_1 + jv_2$, and let $\phi_{i,j}$ denote the integral linear function that agrees with $\|\cdot\|$ on the basis $\{v_1, v_2^i, v_3^{i,j}\}$. Fix a small open set U containing 0 as above, whose closure is contained in the interior of B. For each sufficiently large fixed j, the functions $\phi_{i,j}$ for i big compared to j satisfy $\phi_{i,j} \le 1$ on U. By convexity of B and the discussion above, for each fixed j there is a ψ_j such that $\psi_j = \phi_{i,j}$ for all sufficiently large i (depending on j). The plane $\psi_j = 1$ intersects ∂B in two straight lines joining $v_1/\|v_1\|$ to each of $v_2^i/\|v_2^i\|$ and $v_3^{i,j}/\|v_3^{i,j}\|$. Since $\psi_j \le 1$ on U for each j, there are distinct j, j' for which $\psi_j = \psi_{j'}$. Consequently the plane $\psi_j = 1$ intersects ∂B in three straight lines meeting at acute angles, and therefore (by convexity of B) intersects ∂B in a subset with nonempty interior whose closure contains $v_1/\|v_1\|$.

Since v_1 was arbitrary, we conclude that B is the intersection of the half spaces $\phi \le 1$ where ϕ is integral and linear and satisfies $\phi \le 1$ on B. Since there are only finitely many such ϕ, the lemma follows. □

REMARK 4.2. The proof of Lemma 4.1 is Thurston's proof of the polyhedrality of his norm. Oertel's proof [**162**], using branched surfaces, is closer in spirit to the methods in this chapter, but requires more prerequisites from 3-manifold topology.

There is a similar definition of a norm on $H_2(M)$, defined by restricting attention to closed embedded surfaces representing absolute homology classes. Note that the value of $\|\cdot\|_T$ on any *absolute* class in $H_2(M;\mathbb{Z})$ is an *even* integer.

The crucial property of the Thurston norm, for our purposes, is its relation to the (Gromov) L^1 norm $\|\cdot\|_1$ on $H_2(M, \partial M; \mathbb{R})$. Thurston already showed that a compact leaf of a taut foliation is minimizing in its homology class in both the Thurston and the Gromov norms, and therefore the two norms are proportional on the projective homology classes realized by such surfaces. Conversely, Gabai [**84**] showed that every Thurston norm minimizing surface is a compact leaf of a taut foliation. From this he deduced the following proportionality theorem, conjectured by Thurston:

THEOREM 4.3 (Gabai, Corollary 6.18. [**84**]). *Let M be a compact oriented 3-manifold. Then on $H_2(M)$ or $H_2(M, \partial M)$,*

$$\|\cdot\|_T = \frac{1}{2}\|\cdot\|_1$$

From this we can deduce the following fact:

PROPOSITION 4.4. *Let M be a compact oriented 3-manifold. Let $\gamma \subset \partial M$ be an embedded, oriented loop. Let a be the conjugacy class in $\pi_1(M)$ represented by γ. Suppose $a \in [\pi_1(M), \pi_1(M)]$. Then $\mathrm{scl}(a) \in \mathbb{Q}$. Furthermore, if $H_2(M;\mathbb{R}) = 0$ then $\mathrm{scl}(a) \in \frac{1}{2} + \mathbb{Z}$.*

PROOF. Let A be a regular annulus neighborhood of γ, and let N be obtained by doubling M along A. We write $N = M \cup \overline{M}$ where $M \cap \overline{M} = A$. By Mayer–Vietoris there is an exact sequence

$$0 \to H_2(M) \oplus H_2(\overline{M}) \to H_2(N) \xrightarrow{\partial} H_1(A) \to 0$$

where exactness at the last term follows because the inclusion map of $H_1(A)$ into both $H_1(M)$ and $H_1(\overline{M})$ is zero, because $a \in [\pi_1(M), \pi_1(M)]$. Let $V \subset H_2(N)$ be the integral affine subspace $V = \partial^{-1}([\gamma])$ where $[\gamma] \in H_1(A)$ is the generator. If C

is a 2-chain in M with the support of ∂C mapping into γ, and $[\partial C] = [\gamma]$ in $H_1(A)$, then $C - \overline{C}$ is a 2-cycle in N representing an element of V. It follows that there is an inequality

$$2\,\mathrm{fill}(a) \geq \inf_{v \in V} \|v\|_1$$

Conversely, let S be a Thurston norm minimizing surface in N representing an integral class which is projectively close to an element of V. By making S transverse to A, and isotoping it so that no component of $S \cap M$ or $S \cap \overline{M}$ is a disk, one obtains an inequality

$$\mathrm{scl}(a) \leq \frac{1}{4} \inf_{v \in V} \|v\|_T$$

Using $\mathrm{scl}(a) = \frac{1}{4}\mathrm{fill}(a)$ one therefore obtains an equality

$$\mathrm{scl}(a) = \frac{1}{4} \inf_{v \in V} \|v\|_T$$

Since the Thurston norm takes even integral values on integer lattice points, and since V is an integral affine subspace, the infimum is rational.

In the special case that $H_2(M;\mathbb{R}) = 0$, the subspace V is 0 dimensional, and consists of a single integral class v. If S is a norm minimizing surface representing v, make S transverse to A and efficient. If S_1 and S_2 are the intersections $S_1 \cap M_1$ and $S_2 \cap M_2$ then $\chi(S_1) = \chi^-(S_1) = \chi^-(S_2) = \chi(S_2)$ or else by replacing S_1 by $\overline{S}_2$ (for example) one could reduce the norm. Since each S_i is embedded, the intersection $S_1 \cap A$ consists of a union of embedded loops. Norm minimizing surfaces are incompressible, so each oriented boundary component of S_1 is isotopic in A to γ or γ^{-1}. Moreover by the definition of ∂, there is an equality $[\partial S_1] = [\gamma]$ in $H_1(A)$. It follows that S_1 has an *odd* number of boundary components, and therefore $\|v\|_T = 4n + 2$ for some integer n. Consequently in this case we have an equality

$$\mathrm{scl}(a) = \frac{1}{4}\|v\|_T \in \frac{1}{2} + \mathbb{Z}$$

□

EXAMPLE 4.5. A word w in a free group F is *geometric* if there is a handlebody H with $\pi_1(H) = F$ such that a loop γ in H in the conjugacy class of w is homotopic to an embedded loop in ∂H. For such a w, one has $\mathrm{scl}(w) \in \frac{1}{2} + \mathbb{Z}$ (if $w \in [F, F]$).

A word w in F is *virtually geometric* if there is a finite cover $H' \to H$ such that the total preimage of γ in H' is homotopic to a union of embedded loops in $\partial H'$. If w is virtually geometric, then $\mathrm{scl}(w) \in \mathbb{Q}$.

EXAMPLE 4.6 (Gordon–Wilton [**94**]). In $F_2 = \langle a, b\rangle$, the Baumslag–Solitar words $w = b^{-1}a^p b a^q$ are virtually geometric (but not geometric).

EXAMPLE 4.7 (Manning [**144**]). Jason Manning gives a criterion to show that certain words in free groups are not virtually geometric. For example, in $F_3 = \langle a, b, c\rangle$, many words, including $b^2a^2c^2abc$ and $ba^2bc^2a^{-1}c^{-1}b^{-2}c^{-1}a^{-1}$, are not virtually geometric. Similar examples exist in nonabelian free groups of any rank.

A corollary of Theorem 4.3 is that the unit ball of the dual Thurston norm is the convex hull of the set of cohomology classes which are in the image of elements of H_b^2 whose (L^∞) norm is equal to 1/2. It is natural to try to find explicit bounded 2-cocycles whose cohomology classes correspond to the vertices of the dual norm, and which therefore can be used to certify that a given surface is Thurston norm minimizing. It is a highly nontrivial fact that for every irreducible, atoroidal 3-manifold, one may find a finite collection of classes $[e] \in H^2$ in the image of H^2_b, whose convex hull is equal to the unit ball of the dual norm, and such that every $[e]$ is obtained by pulling back the Euler class (i.e. the generator of H^2) from the group $\text{Homeo}^+(S^1)$ under some faithful homomorphism $\pi_1(M) \to \text{Homeo}^+(S^1)$.

In fact, this characterization of the Thurston norm is unfamiliar even to many people working in 3-manifold topology, and deserves some explanation. A homomorphism $\pi_1(M) \to \text{Homeo}^+(S^1)$ is the same thing as an action of $\pi_1(M)$ on a circle. Gabai's main theorem from [**84**] (Theorem 5.5) says that every embedded surface S realizing $\|\cdot\|_T$ in its homology class is a leaf of a finite depth taut foliation $\mathcal{F}$ on M. To every taut foliation of an atoroidal 3-manifold one can associate a *universal circle* S^1_{univ}, which monotonely parameterizes the circle at infinity of every leaf of $\tilde{\mathcal{F}}$, the pullback of $\mathcal{F}$ to the universal cover $\tilde{M}$. See [**40**] Chapter 7 for a proof, and an extensive discussion of universal circles. The construction of S^1_{univ} is natural, so the action of $\pi_1(M)$ as the deck group of $\tilde{M}$ induces an action on S^1_{univ} by homeomorphisms, and therefore a representation $\rho_{\text{univ}} : \pi_1(M) \to \text{Homeo}^+(S^1_{\text{univ}})$. Associated to this representation there is a foliated circle bundle E over M, which one can show is isomorphic (as a circle bundle) to the unit tangent bundle to the foliation $UT\mathcal{F}$. In particular, the pullback $[e]$ of the Euler class is the obstruction to finding a section of $UT\mathcal{F}$, and $[e](S) = \pm\chi(S)$ by construction. On the other hand, the Milnor–Wood inequality (Theorem 2.52) implies that $\|[e]\|_\infty = 1/2$, so this class is in the boundary of the convex hull of the dual norm.

In light of this fact, it is natural to wonder whether the unit ball of the scl norm on $B^H_1(\pi_1(M))$ for M an irreducible, atoroidal 3-manifold is cut out by hyperplanes determined by rotation quasimorphisms. In fact, it turns out that this is *not* the case. A counterexample is the Weeks manifold W, which can be obtained by $(5/1, 5/2)$ surgery on the components of the Whitehead link in S^3, and is known (see e.g. Milley [**153**]) to be the smallest volume closed orientable hyperbolic 3-manifold. In [**48**] it is shown that every homomorphism $\pi_1(W) \to \text{Homeo}^+(S^1)$ must factor through $\mathbb{Z}/5\mathbb{Z}$, and therefore there are *no* nontrivial rotation quasimorphisms on $\pi_1(W)$. To reconcile this with the assertions in the previous paragraph, note that W is a rational homology sphere, so $H_2(W)$ is trivial.

It should be clear from this example that the relationship between the scl norm and rotation quasimorphisms (at least in 3-manifold groups) cannot be as straightforward as one might naively guess, based on familiarity with the Thurston norm (also, see Example 4.35). Thus, although in the next few sections we give a direct proof that the scl norm on free groups is piecewise rational linear, our argument does not suggest a natural family of extremal quasimorphisms which define the faces of the unit ball (however, see § 4.2).

4.1.2. Branched surfaces. It is convenient to introduce the language of branched surfaces. For a reference, see [**160**] or § 6.3 of [**40**].

DEFINITION 4.8. A *branched surface* B is a finite, smooth 2-complex obtained from a finite collection of smooth surfaces by identifying compact subsurfaces.

The *branch locus* of B, denoted $\text{br}(B)$, is the set of points which are not 2-manifold points. The components of $B - \text{br}(B)$ are called the *sectors* of the branched surface. The set of sectors of B is denoted $S(B)$. A *simple branched surface* is a branched surface for which the branch locus is a finite union of disjoint smoothly embedded simple loops and simple proper arcs.

In a simple branched surface, local sectors meet along segments of the branch locus. The local sheets approach the branch segment from one of two sides (distinguished by the smooth structure along $\mathrm{br}(B)$). In a generic branched surface, three sheets meet along each component of the branch locus, two on one side and one on the other. However, the branched surfaces we consider in this section are *not* generic, and any (positive) number of sheets may meet a segment of branch locus on either side. See Figure 4.1 for an example.

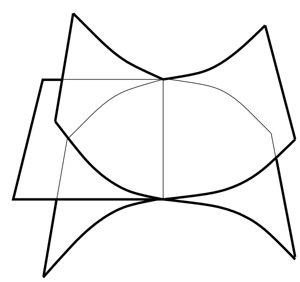


FIGURE 4.1. An example of a local model for a simple branched surface. In this example, five sheets meet along the branch locus, two on one side and three on the other.

Branched surfaces can have boundary or not. The branched surfaces considered in this section have boundary. We require that the branched locus intersect the boundary transversely. Note that the sectors of a simple branched surface B are themselves surfaces, perhaps with boundary, and possibly with *corners* where arcs of the branch locus intersect ∂B. A branched surface B is oriented or not according to whether the sectors can be compatibly oriented. We are exclusively interested in oriented branched surfaces.

DEFINITION 4.9. Let B be a simple branched surface. A *weight* on B is a function $w : S(B) \to \mathbb{R}$ such that for each component γ of $\mathrm{br}(B)$, the sum of the values of w on the sectors which meet γ on one side is equal to the sum of the values of w on the sectors which meet γ on the other side. A weight is *rational* if it takes values in $\mathbb{Q}$, and *integral* if it takes values in $\mathbb{Z}$.

It follows from Definition 4.9 that the set of weights on B is a subspace of $\mathbb{R}^{|S(B)|}$ defined by a finite family of integral linear equalities, one equality for each component of $\mathrm{br}(B)$.

NOTATION 4.10. Let $W(B)$ denote the (finite dimensional) real vector space of weights on B, and $W^+(B)$ the convex cone of weights which take non-negative values on every sector. If B is understood, abbreviate these spaces by W and W^+.

There is a close relationship between (non-negative integral) weights on a branched surface B and surfaces mapping to B in a particularly simple way. Since B has a smooth structure, it makes sense to say that a map $f : S \to B$ is an immersion, when S is a smooth surface.

DEFINITION 4.11. Let B be an oriented simple branched surface, possibly with boundary. A *carrying map* is a proper, orientation-preserving immersion $f : S \to B$ from some compact oriented surface S (possibly with boundary) to B. By abuse of notation we say that B *carries* S.

A carrying map $f : S \to B$ determines a non-negative integral weight $w(f)$, whose value on each sector $\sigma \in S(B)$ is the local degree of f along σ. Since a carrying map is an orientation-preserving immersion, the local degree along a sector σ is equal to the number of preimages of any point in the interior. In other words,

$$w(f)(\sigma) = \#\{f^{-1}(p)\} \text{ for } p \in \sigma$$

LEMMA 4.12. *Let B be a simple branched surface. Every non-negative integral weight on B is represented by a carrying map. Conversely, if $f : S \to B$ represents a weight w, then $\chi(S)$ depends only on w, and is a rational linear function of the co-ordinates of $w \in W$.*

PROOF. Let w be a non-negative integral weight. For each sector $\sigma \in S(B)$, take $w(\sigma)$ copies of σ. At each $\gamma \in \mathrm{br}(B)$, the sum of the weights on the sectors on one side is equal to the sum of the weights on sectors on the other side. Choose a bijection between the two sets of copies of sectors, and glue the copies according to this bijection along their edges corresponding to γ. The result of this gluing is a surface S, which comes together with a tautological orientation-preserving immersion to B, realizing the weight w. Moreover, all surfaces representing w arise this way, for various choices of bijections as above.

Each sector $\sigma \in S(B)$ can be thought of as a surface with corners. The corners are the points where arcs of $\mathrm{br}(B)$ run into ∂B. Each such surface σ has an *orbifold* Euler characteristic $\chi_o(\sigma)$ defined by the formula

$$\chi_o(\sigma) = \chi(\sigma) - c(\sigma)/4$$

where $\chi(\cdot)$ denotes ordinary Euler characteristic of the underlying surface, and $c(\cdot)$ denotes the number of (boundary) corners. If a smooth surface S is obtained by gluing surfaces S_i with corners, then $\chi(S) = \sum_i \chi_o(S_i)$. Hence if S is a surface with weight w, then $\chi(S) = \sum_\sigma w(\sigma)\chi_o(\sigma)$, which depends only on w, as claimed. □

REMARK 4.13. Lemma 4.12, though simple to state and prove, is actually surprisingly delicate. The reader whose intuitions have been honed by exposure to train-tracks in surfaces, or *embedded* branched surfaces in 3-manifolds, may not appreciate how subtle such objects really are.

In great generality, a compact Riemann surface lamination is carried by an abstract branched surface, and the space of weights on such a surface is finite dimensional (see [**160**]). For a branched surface embedded in a 3-manifold, a non-negative integral weight determines a unique *embedded* surface which maps to the branched surface by an immersion. However the construction of such a surface depends on the local transverse order structure on branches inherited by codimension 1 objects in a 3-manifold.

If B is an abstract (not necessarily simple) branched surface, and w a non-negative integral weight on B, then from w one can construct a surface S mapping to B, but the map is in general a *branched* immersion, branched over the vertices of $\mathrm{br}(B)$, and χ depends not only on w but on the way S branches at each such point.

To associate an (unbranched) carrying map to a weight one must solve a holonomy problem. Moreover, it might be the case that this holonomy problem can be solved for nw but not for w, where w is a non-negative integral weight, and n is a positive integer.

A similar, and completely analogous phenomenon occurs when one tries to do *immersed* normal surface theory in 3-manifolds.

By contrast, the function χ^- might well depend on the choice of a surface S representing a weight w. For, the number of disk components of S might depend on the way in which sectors are glued up. This motivates the following definition.

DEFINITION 4.14. An oriented simple branched surface is *essential* if it does not carry a disk or sphere.

EXAMPLE 4.15. If every sector satisfies $\chi_o(\sigma) \le 0$ then $\chi(S) \le 0$ for any surface carried by B. Consequently in this case, B is essential.

If S is carried by an essential simple branched surface, then every component of S has non-positive Euler characteristic. Consequently $\chi(S) = \chi^-(S)$, and therefore we obtain the following corollary:

COROLLARY 4.16. *Let B be an essential simple branched surface. Then $-\chi^-(S)$ is a linear function of w, where S is a surface realizing a (non-negative integral) weight w.*

4.1.3. Alternating words. As a warm-up, we prove rationality of scl on certain special elements in the free group of rank 2, where the argument is especially transparent. Throughout the sequel we fix notation $F = \langle a, b\rangle$.

DEFINITION 4.17. A word $w \in F$ is *alternating* if it has even length, and the letters alternate between one of $a^{\pm 1}$ and one of $b^{\pm 1}$.

Every alternating word is cyclically reduced. An alternating word is in $[F, F]$ if there are the same number of a's as a^{-1}'s, and similarly for b's and b^{-1}'s. Hence an alternating word in $[F, F]$ has length divisible by 4.

EXAMPLE 4.18. $aba^{-1}b^{-1}$ and $aba^{-1}b^{-1}a^{-1}bab^{-1}$ are examples of alternating words in $[F, F]$.

EXAMPLE 4.19. A word is alternating if and only if in the graphical calculus (see § 2.2.4) it is represented by a loop without backtracks in which every straight segment has length 1.

In what follows, let H be a handlebody of genus 2. We think of H as the union of two solid handles H^+, H^-, glued along a disk E which we call the *splitting disk*. For psychological convenience, we think of H embedded in $\mathbb{R}^3$ in such a way that E is horizontal, H^+ is above, and H^- is below. Let $D^\pm$ be compressing disks for the meridians of $H^\pm$; psychologically, we think of these disks as vertical.

Identify $\pi_1(H)$ with F in such a way that b is represented by the core of the handle H^+ and a is represented by the core of the handle H^-. An alternating word is represented by a particularly simple free homotopy class of loop in H, namely as a union of arcs from E to itself which wind once around either H^+ or H^-, crossing D^+ or D^- transversely in a single point; say that such a representative is in *bridge position*. By convention we assume that a loop in bridge position is *embedded* in H. This is mainly for psychological rather than logical convenience; the isotopy class of γ in H is not relevant in the sequel, only its homotopy class.

In what follows, fix an alternating word w and let γ be a corresponding loop in H in bridge position. Without loss of generality, we can write

$$w = a^{e_1}b^{f_1}a^{e_2}b^{f_2}\cdots a^{e_m}b^{f_m}$$

where each e_i, f_i is ± 1, and m is even, and equal to half the word length of w. Then γ is a union of arcs

$$\gamma = \alpha_1 \cup \beta_1 \cup \cdots \cup \beta_m$$

where α_i is properly embedded in H^-, and winds in the positive direction if $e_1 = 1$, and the negative direction otherwise, and β_i is properly embedded in H^+ and winds similarly according to the sign of f_i. Note that the α_i, β_i are oriented arcs, and the end point of α_i is equal to the initial point of β_i for each i, while the end point of β_i is equal to the initial point of α_{i+1} (indices taken cyclically) for each i.

Let $f : S, \partial S \to H, \gamma$ satisfy $f_*[\partial S] = n[\gamma]$. Recall, by Proposition 2.10 that $\text{scl}(w)$ is equal to the infimum of $-\chi^-(S)/2n(S)$ over all such surfaces. We will show that after possibly replacing S with a simpler surface S' with $n(S') = n(S)$ and $-\chi^-(S') < -\chi^-(S)$, we can homotope f into a particularly simple form.

Assume without loss of generality that S has no disks or closed components, or simple compressing loops, or else $-\chi^-$ could be reduced without affecting n. If some boundary component of S maps to γ with degree 0, we can compress it, reducing $-\chi^-$. So assume that every boundary component maps with nonzero degree, and homotope f so that the restriction of f to each component of ∂S is a covering map to γ. Then perturb f rel. boundary to an immersion in general position with respect to $D^\pm$.

After this perturbation, the preimage $f^{-1}(D^+) \cap S$ is a union of *disjoint, embedded* proper arcs and loops in S. Since by hypothesis S has no simple compressing loops, all the loops are inessential in S, and can be pushed off D^+ by a homotopy of f. Since the restriction of f to ∂S is a covering map, there are no inessential arcs in $f^{-1}(D^+)$, so we may assume that $f^{-1}(D^+)$ consists of a union of disjoint essential embedded proper arcs in S. Do the same for $f^{-1}(D^-)$. After this modification, $f^{-1}(D^+ \cup D^-)$ is a union δ of disjoint essential embedded proper arcs. Let $\mathcal{R}$ be a union of (relatively) open regular neighborhoods in S of the components of δ. The components of $\mathcal{R}$ are called *rectangles*.

The complement of tubular neighborhoods of the $D^\pm$ in H deformation retracts down to the splitting disk E. In fact, there is a deformation retraction of pairs

$$H - N(D^\pm), \gamma - (\gamma \cap N(D^\pm)) \to E, E \cap \gamma$$

Drag f by this deformation retraction, so that after a homotopy, $\mathcal{R}$ is exactly equal to $f^{-1}(H - E)$.

Now consider the components of $S - \mathcal{R}$. Each such component P is a compact surface, whose boundary is broken up into vertices (points in ∂S in the closure of a rectangle of $\mathcal{R}$) and two different kinds of edges: components of $P \cap \partial S$, and components of P in the closure of a rectangle. We refer to the first kind of edges as *boundary edges* and the second kind as *branch edges*. After the homotopy, each boundary edge maps by f to a single point of $\gamma \cap E$, and each branch edge maps to an arc in E. Since E is a disk, if P is not a disk, it contains an essential embedded loop which maps to a null-homotopic loop in E and can therefore be compressed. Since by hypothesis S contains no simple compressing loops, every component P of $S - \mathcal{R}$ is topologically a disk. Since its boundary has a natural cellulation into edges and vertices, we think of P as a *polygon*, whose edges alternate between boundary edges and branch edges. Let $\mathcal{P}$ denote the union of these polygons, and let P_i denote a typical polygon.

For each P_i, let $|P_i|$ denote the number of branch edges of P_i. Observe that the branch edges alternate between arcs bounding rectangles mapping to H^+ and rectangles mapping to H^-. Consequently, each P_i has an even number of branch edges; denote this number by $|P_i|$. Say that a branch edge of P_i *faces up* if it bounds a rectangle mapping to H^+, and it *faces down* otherwise. There are twice as many corners of P_i as branch edges, hence $2|P_i|$ corners.

Since each P_i is topologically a disk, we can compute $\chi_o(P_i) = 1 - |P_i|/2 \le 0$. Similarly, each rectangle of $\mathcal{R}$ has $\chi_o = 0$. Hence

$$-\chi^-(S) = -\chi(S) = \sum_i \frac{|P_i| - 2}{2}$$

Now fix a single polygon P_i. Suppose that there is a point p of $\gamma \cap E$ and two distinct boundary edges e_1, e_2 of P_i which both map to p. Let β be an embedded arc in P_i joining e_1 to e_2. Doing a boundary compression along β reduces $-\chi^-$ by 1. Hence after repeatedly performing such compressions, we can assume (at the cost of replacing the original surface with another of smaller $-\chi^-$) that every polygon P_i has *at most* $|w|$ boundary edges, which map to *distinct* points of $\gamma \cap E$.

Notice what we have achieved in this discussion. Starting with an arbitrary map $f : S, \partial S \to H, \gamma$ we obtained (after homotopy, compression and boundary compression) a new surface and a new map (which by abuse of notation we still denote S, f) such that S is decomposed into two kinds of pieces: *rectangles* which map over the handles of H, and which run between a pair of arcs of γ, and *polygons* which map to the splitting disk E. Each rectangle is determined, up to homotopy, by the pair of arcs of γ that it runs between. Each polygon is determined up to homotopy by a cyclically ordered list of *distinct* elements of $\gamma \cap E$ that the boundary edges map to in order, and by the data of whether each branch edge faces up or down. There are only finitely many combinatorial possibilities for each rectangle and for each polygon. Thus the surface S is built from finitely many pieces, all drawn from a finite set of combinatorial types.

This last observation is crucial, and reduces the computation of $\mathrm{scl}(w)$ to a *finite* integer linear programming problem. We explain how.

Build an oriented essential simple branched surface B as follows. The sectors of B are the disjoint union of all possible polygons (with boundary edges mapping to distinct points of $E \cap \gamma$) and all possible rectangles. Glue up rectangles to polygons in all possible orientation-preserving ways, ensuring that branch edges that face up and down are only glued to rectangles in H^+ and H^- respectively. The result is an abstract branched surface B and a homotopy class of map $\iota : B \to H$ taking ∂B to γ.

There are two components of the branch locus for each pair of distinct points in $E \cap \gamma$, distinguished by whether such components bound rectangles in H^+ or in H^-. In particular, the branch locus is a 1-manifold, and therefore the branched surface is simple. Furthermore, each polygon contributes non-positively to χ_o and each rectangle contributes 0, so the branched surface is essential.

Since every surface $f : S, \partial S \to H, \gamma$ can be compressed, boundary compressed and homotoped until it is made up of rectangles and polygons, we conclude the following:

LEMMA 4.20. *Let B denote the essential simple branched surface, constructed as above. Then every $f : S, \partial S \to H, \gamma$ can be compressed, boundary compressed and homotoped without increasing $-\chi^-$, to a map which is carried by B.*

Notice that the branched surface B can be constructed effectively from the word w. Let $w \in W^+$ be a non-negative integral weight on B. Let $f : S \to B$ be a carrying map with weight w. The composition $\iota \circ f : S \to H$ takes $\partial S \to \gamma$. Define $\partial(w) = n(S)$, and extend by linearity and continuity to a rational linear map $\partial : W^+ \to \mathbb{R}$. By construction,

$$\mathrm{scl}(w) = \inf_{w \in W^+ \cap \partial^{-1}(1)} \frac{-\chi^-(w)}{2}$$

But $W^+ \cap \partial^{-1}(1)$ is a closed rational polyhedron, and $-\chi^-$ is a rational linear function which is non-negative on the cone W^+, and therefore achieves its infimum on a closed rational polyhedron Q in $W^+ \cap \partial^{-1}(1)$. It follows that $\mathrm{scl}(w)$ is rational. Moreover, given W^+ and the functions $-\chi^-$ and ∂, computing the polyhedron Q is a finite linear programming problem which can be solved by any one of a number of methods. Thus there is an effective algorithm to compute $\mathrm{scl}(w)$.

4.1.4. Bridge position. We extend the arguments in § 4.1.3 in several ways: to free groups of arbitrary rank, and to arbitrary finite integral linear combinations of arbitrary elements.

Let F be a free group with generators a_i. For each i, let H_i denote a solid torus with a marked disk E_i in its boundary, and let H be obtained from the H_i by identifying the E_i with a single disk E. If the rank of F is 2, this is an ordinary genus 2 handlebody, and H_1, H_2 are H^+, H^- from the last section. For each i, let D_i be a decomposing disk for the handlebody H_i, disjoint from E, and denote the union of the D_i by $\mathcal{D}$. Let $w \in F$ be cyclically reduced. The conjugacy class of w determines a free homotopy class of loop in H; we will choose a representative γ in this free homotopy class whose intersection with E and $\mathcal{D}$ is simple.

A *vertical arc* is an arc with endpoints on E whose interior is properly embedded in some $H_i - E$. A *horizontal arc* is an arc embedded in E. The representative γ will have one vertical arc in H_i for each appearance of $a_i^{\pm}$ in w, and one horizontal arc between any two consecutive appearances of $a_i^{\pm}$ (notice, since w is cyclically reduced, that consecutive appearances of $a_i^{\pm}$ have the same sign). This uniquely determines the homotopy class of γ.

DEFINITION 4.21. A representative γ in the free homotopy class corresponding to the conjugacy class of w, constructed as above, is said to be in *bridge position.*

REMARK 4.22. For rank 2 and for alternating words, this agrees with the definition from § 4.1.3.

Let $w_1, \cdots, w_n$ be a finite collection of elements which are cyclically reduced in their conjugacy class, and $\gamma_1, \cdots, \gamma_n$ loops in bridge position in H. Denote the union of the γ_i by Γ. Let $f : S, \partial S \to H, \Gamma$ be given, and assume that S has no disk or closed components, or simple compressing loops. As in § 4.1.3, after a homotopy we can assume that $f^{-1}(\mathcal{D})$ is a union of disjoint essential embedded proper arcs, and $\mathcal{R} = f^{-1}(H - E)$ is a union of disjoint embedded rectangles with the components of $f^{-1}(\mathcal{D})$ as their cores. Since S has no simple compressing loops, as in § 4.1.3 we can conclude that every component P_i of $S - \mathcal{R}$ is a polygon.

The branch edges of the P_i are edges in the closure of components of $\mathcal{R}$, but there are two kinds of boundary edges: those which map to a single endpoint of a vertical arc of some γ_i, and those which map to a horizontal edge. As before, if some polygon has boundary edges e_i, e_j mapping to the same point or horizontal arc of $E \cap \Gamma$, we can do a boundary compression of S to reduce $-\chi^-$. So without loss of generality, we conclude that distinct boundary edges e_i, e_j of the same polygon map to different points or arcs of $E \cap \Gamma$.

Let $|P_i|$ denote the number of branch edges of P_i. As a surface with corners, we have $c(P_i) = 2|P_i|$ so $\chi_o(P_i) = 1 - |P_i|/2$. Rectangles contribute 0 to χ_o, so

$$-\chi^-(S) = -\chi(S) = \sum_i \frac{|P_i| - 2}{2}$$

One can build a simple essential branched surface B as before, together with a homotopy class of map $\iota : B \to H$ with $\iota(\partial B) = \Gamma$. Every map $f : S, \partial S \to H, \Gamma$ can be compressed, boundary compressed and homotoped until it factors through a carrying map to B.

Let K be ker : $H_1(\Gamma) \to H_1(H)$ induced by inclusion. The vector space K is isomorphic to the intersection $B_1(F) \cap \langle w_1, \cdots, w_n \rangle$. The inclusion map on homology is defined over $\mathbb{Z}$, so K is a rational subspace of $H_1(\Gamma)$. With notation as in § 4.1.3, there is a surjective rational linear map $\partial : W^+ \to K$. For each $k \in K$ there is an equality

$$\text{scl}(k) = \inf_{w \in W^+ \cap \partial^{-1}(k)} \frac{-\chi^-(w)}{2}$$

Now, W^+ is a finite dimensional rational convex polyhedron with finitely many extremal rays, each passing through a rational point v_i, and $-\chi^-$ is a rational linear function. Therefore

$$\text{scl}(k) = \inf \frac{\sum_i -t_i \chi^-(v_i)}{2}$$

where the infimum is taken over all non-negative t_i for which $\sum_i t_i \partial(v_i) = k$. Explicitly, each basis $\mathcal{S}$ of elements v_i determines a rational linear function $f_{\mathcal{S}}$ on W^+ whose value is $-\chi^-(v_i)/2$ on $v_i \in \mathcal{S}$, and $\text{scl} \circ \partial$ is the minimum of this finite collection of functions. In other words, $\text{scl} \circ \partial$ is a piecewise rational linear function on W^+ and therefore scl is piecewise rational linear on K.

Recall that a map $f : S \to H$ is *extremal* if it realizes the infimum, over all surfaces without closed or disk components, of $-\chi^-(S)/2n(S)$. If w is a non-negative rational weight realizing the infimum of $-\chi^-(w)/2$ on $\partial^{-1}(k)$ for some rational class $k \in B_1^H(F)$, then some integral multiple of w is integral. Any carrying map realizing this weight gives rise to an extremal surface, and all extremal surfaces arise in this way.

We have now completed the proof of the Rationality Theorem. In order to state the theorem precisely, we must first say what we mean for a function on an infinite dimensional vector space to be *piecewise rational linear*.

Definition 4.23. Let V be a real vector space. A function ϕ on V is *piecewise linear* if for every finite dimensional subspace W of V, the restriction of ϕ to W is piecewise linear. If $V = V_{\mathbb{Q}} \otimes \mathbb{R}$ where $V_{\mathbb{Q}}$ is a (given) rational vector space, a subspace $W \subset V$ is *rational* if it is of the form $W = W_{\mathbb{Q}} \otimes \mathbb{R}$ for some subspace $W_{\mathbb{Q}} = V_{\mathbb{Q}} \cap W$. A function ϕ on V is *piecewise rational linear* if for every finite dimensional rational subspace W of V, the restriction of ϕ to W is piecewise linear, and rational on $W_{\mathbb{Q}}$.

Recall from § 2.6.2 that for any group G, the space $B_1(G)$ is the vector space of real (group) 1-boundaries, and $B_1^H(G)$ is the quotient of $B_1(G)$ by the subspace H spanned by elements of the form $g^n - ng$ and $g - hgh^{-1}$. In general, scl is a pseudo-norm on B_1^H, but when G is hyperbolic, scl is a genuine norm (Corollary 3.57).

The results of this section prove the following theorem.

Theorem 4.24 (Rationality Theorem). *Let F be a free group.*

(1) $\mathrm{scl}(g) \in \mathbb{Q}$ *for all* $g \in [F, F]$.
(2) *Every* $g \in [F, F]$ *bounds an extremal surface.*
(3) *The function* scl *is a piecewise rational linear norm on* $B_1^H(F)$.
(4) *Every nonzero finite rational linear chain* $A \in B_1^H(F)$ *projectively bounds an extremal surface*
(5) *There is an algorithm to calculate* scl *on any finite dimensional rational subspace of* $B_1^H(F)$*, and to construct all extremal surfaces in a given projective class.*

Remark 4.25. Note by Proposition 2.104 that every extremal surface as above is π_1-injective.

4.1.5. PQL groups. Motivated by the results of the previous section, we define the following class of groups.

Definition 4.26. A group G is PQL (pronounced "pickle") if scl is piecewise rational linear on $B_1^H(G)$.

Example 4.27. An amenable group is trivially PQL, by Theorem 2.47 and Theorem 2.79.

Example 4.28. Theorem 4.24 implies that finitely generated free groups are PQL. Suppose F is an infinitely generated free group. Since any finite subset of $B_1^H(F)$ is contained in the image of $B_1^H(F_n)$ for some finitely generated summand F_n, we conclude that F is also PQL.

There are a few basic methods to derive new PQL groups from old.

Proposition 4.29. *Let H be a subgroup of G of finite index. Then if H is PQL, so is G.*

Proof. Let X be a space with $\pi_1(X) = G$. Let $g_1, \cdots, g_m$ be elements of G whose conjugacy classes are represented by loops $\gamma_1, \cdots, \gamma_m$. Let $\widehat{X}$ be a finite cover of X with $\pi_1(\widehat{X}) = H$. For each i, let $\beta_{i,j}$ be the preimages of γ_i in $\widehat{X}$, and let $h_{i,j} \in H$ be elements whose conjugacy classes represent the $\beta_{i,j}$. By Proposition 2.80, for any integers $n_1, \cdots, n_m$ we have

$$\mathrm{scl}_G(\sum_i n_i g_i) = \frac{1}{[G:H]} \cdot \mathrm{scl}_H(\sum_{i,j} n_i h_{i,j})$$

and the proposition follows. □

Hence virtually free groups are PQL. This class of groups includes fundamental groups of non-compact hyperbolic orbifolds.

Proposition 4.30. *Let $A \xrightarrow{i} G \xrightarrow{q} H \to 1$ be an exact sequence, where A is amenable and H is PQL and satisfies $H^2(H;\mathbb{R}) = 0$. Then G is PQL.*

Proof. Since A is amenable, Theorem 2.47 says that the bounded cohomology of A vanishes in each dimension. By Theorem 2.50 and Theorem 2.49 one obtains a commutative diagram as in Figure 4.2 with exact rows and columns. Let $\alpha \in Q(G)$

$$\begin{array}{ccccccc}
H^1(H) & \longrightarrow & Q(H) & \xrightarrow{\delta} & H_b^2(H) & \longrightarrow & 0 \\
\downarrow & & \downarrow{\scriptstyle q^*} & & \downarrow{\scriptstyle q^*} & & \\
H^1(G) & \longrightarrow & Q(G) & \xrightarrow{\delta} & H_b^2(G) & \longrightarrow & H^2(G) \\
\downarrow & & \downarrow & & \downarrow & & \\
H^1(A) & \longrightarrow & Q(A) & \longrightarrow & 0 & &
\end{array}$$

Figure 4.2. This diagram has exact columns (by Theorem 2.49) and exact rows (by Theorem 2.50).

be given. Then $\delta\alpha \in H_b^2(G)$ is equal to $q^*\beta$ for some $\beta \in H_b^2(H)$, since $H_b^2(H) \to H_b^2(G)$ is surjective. Since $H^2(H)$ is zero, there is some $\gamma \in Q(H)$ with $\delta\gamma = \beta$, and therefore $\alpha - q^*\gamma \in Q(G)$ is in the image of $H^1(G)$. Since α was arbitrary, this says that the composition $Q(H) \to Q(G) \to Q(G)/H^1(G)$ is surjective.

It is a general fact that for any surjection of groups $q : G \to H$, and any quasimorphism ϕ on H, there is an equality $D(\phi) = D(q^*\phi)$ where the left side is the defect of ϕ on H, and the right side is the defect of $q^*\phi$ on G. For,

$$D(q^*\phi) = \sup_{a,b,\in G} |\phi(q(a)) + \phi(q(b)) - \phi(q(ab))| = D(\phi)$$

where the second equality follows from the definition of $D(\phi)$ and surjectivity. By Theorem 2.79, for any $\sum t_i a_i \in B_1^H(G)$ we have

$$\begin{aligned}
\mathrm{scl}_G\left(\sum t_i a_i\right) &= \frac{1}{2} \sup_{\phi \in Q(G)/H^1(G)} \frac{\sum_i t_i \phi(a_i)}{D(\phi)} \\
&= \frac{1}{2} \sup_{\phi \in Q(H)/H^1(H)} \frac{\sum_i t_i q^*\phi(a_i)}{D(q^*\phi)} \\
&= \frac{1}{2} \sup_{\phi \in Q(H)/H^1(H)} \frac{\sum_i t_i \phi(q(a_i))}{D(\phi)} \\
&= \mathrm{scl}_H\left(\sum t_i q(a_i)\right)
\end{aligned}$$

It follows that G is PQL if H is, as claimed. □

Remark 4.31. If $H^2(H)$ is nonzero, there might be elements in $Q(G)/H^1(G)$ which are not in the image of $Q(H)$. If H is finitely presented, $H^2(H)$ is finitely generated, so $Q(G)/(H^1(G) + q^*Q(H))$ is finite dimensional and is generated by a finite number of quasimorphisms $\phi_1, \cdots, \phi_n$. If one can find generators ϕ_i as above which take on rational values on rational elements of $B_1^H(G)$, then if H is PQL, so is G.

Corollary 4.32. *Let M be a noncompact Seifert-fibered 3-manifold. Then $\pi_1(M)$ is PQL.*

Proof. For M as above there is a central extension $\mathbb{Z} \to \pi_1(M) \to G$ where G is the fundamental group of a noncompact surface orbifold. If G is amenable, so is $\pi_1(M)$, and $\pi_1(M)$ is trivially PQL. Otherwise G is virtually free. In this case there is a finite index subgroup H of $\pi_1(M)$ which is a product $\mathbb{Z} \oplus F$ where F is free. By Proposition 4.30, the group H is PQL, and therefore by Proposition 4.29, so is $\pi_1(M)$. □

Example 4.33. Let M be homeomorphic to $S^3 - K$ where K is the trefoil knot. Then M is Seifert fibered and noncompact, so $\pi_1(M)$ is PQL. It is well-known that $\pi_1(M)$ is isomorphic to the braid group B_3 (see e.g. [**16**]).

4.1.6. Implementing the Algorithm. In this section we discuss in more explicit terms the algorithm described implicitly in the last few sections. Proposition 2.13 implies that we can restrict attention to *monotone* admissible maps in order to calculate scl. If $f : (S, \partial S) \to (H, \gamma)$ is monotone, the restriction $\partial S \to \gamma$ is orientation-preserving. This reduces the number of rectangle types that must be considered by roughly a factor of 4, and concomitantly reduces the number of polygon types.

We show how the algorithm runs in practice. For convenience, we restrict attention to alternating words in F_2. In what follows, for the sake of legibility, we denote a^{-1} by A and b^{-1} by B.

Example 4.34. Let $w = abABAbaB$. The loop γ is a union of 8 arcs, each arc corresponding to a letter in w. The initial vertex of each arc is a point on E; denote these points $v_0, v_1, \cdots, v_7$. An *admissible arc* is an arc that might be contained in a polygon in a monotone extremal surface. Such an arc is given by an ordered pair (v_i, v_j) where v_i is the initial vertex of an arc corresponding to some letter x or X and v_j is the terminal vertex of an arc corresponding to a letter X or x. Since w is alternating, there are $|w|/4 = 2$ copies of each of the letters a, A, b, B and consequently there are $|w|^2/4 = 16$ admissible arcs (the arc (v_i, v_j) is denoted ij for brevity):

$$03, 21, 14, 32, 05, 41, 10, 72, 27, 63, 54, 36, 47, 65, 50, 76$$

A *polygon* is a cyclically ordered list of vertices, where no vertex appears more than once, and each consecutive pair of vertices is an admissible arc. There are 18 polygons:

$$03214765, 0321, 03276541, 032765, 036541, 03654721, 0365, 2147, 214763,$$

$$210547, 21054763, 14, 3276, 0541, 05, 72, 63, 5476$$

(note that each polygon has an even number of vertices). Each rectangle bounds two admissible arcs, but there is a relation between these two arcs: if a rectangle bounds ij at one end, it bounds $(j-1)(i+1)$ at the other end. The linear programming problem takes place in the vector space $P \cong \mathbb{R}^{18}$ spanned by a basis p_i whose co-ordinates count the number of polygons of type i. Each rectangle imposes one equation, of the form $\sum p_k = \sum p_l$ where the left hand side counts the number of polygons that contain an admissible edge ij and the right hand side counts the number of polygons that contain an admissible edge $(j-1)(i+1)$ (note that a polygon type might contain both or neither). There are twice as many admissible edges as equations, and hence $|w|^2/8 = 8$ equations:

$$p_0 + p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = p_0 + p_1 + p_5 + p_7 + p_8 + p_9 + p_{10}$$

$$p_0 + p_7 + p_8 + p_{11} = p_0 + p_1 + p_2 + p_3 + p_8 + p_{10} + p_{12}$$

and so on.

Restricting to geometrically sensible answers imposes the conditions that each $p_i \ge 0$. For each i, let l_i denote the number of branch edges in the polygon of type i. In this example, l_i is equal to the length of the corresponding string of vertices; hence $l_0 = 8$, $l_1 = 4$, $l_2 = 8$, $l_3 = 6$ and so on. To normalize the solution so that the boundary represents $[\gamma]$ in homology, we need to impose the equation

$$\sum_i l_i p_i = |w| = 8$$

Subject to this list of constraints, $\text{scl}(w)$ is the minimum of the *objective function*

$$\frac{-\chi^-}{2} = \sum_i \frac{(l_i - 2)p_i}{4}$$

This linear programming problem can be solved using exact arithmetic, for instance using the GNU package `glpsol` ([**140**]) and Masashi Kiyomi's program `exlp` ([**128**]), returning the answer $\text{scl}(w) = 0.5$. Moreover, an extremal solution describes how to construct an extremal surface consisting of one 4-gon and two bigons $0541 + 72 + 63$ and four rectangles. This exhibits γ as the boundary of a once-punctured torus, and shows that w is a commutator (which is easily seen in any case: $abABAbaB = [a, bAB]$).

See e.g. Dantzig [**62**] for an introduction to linear programming.

Example 4.35. Bavard [**8**] p. 148 asked whether scl in the commutator subgroup of free group takes on values in $\frac{1}{2}\mathbb{Z}$. This should be viewed in some sense as the natural analogue of the fact that in a 3-manifold M, the (Gromov-)Thurston norm takes on values in $2\mathbb{Z}$ on the integral lattice $H_2(M;\mathbb{Z})$ (also compare with Proposition 4.4). In fact, the answer to Bavard's question is negative: there are many elements in free groups whose scl is not a half-integer. One explicit example is $w = baBABAbaBabA$; the identity

$$[abaB, ABAbaBabAbaBABAbaBababABB] \cdot [ABAba, BabAbaBABAbba] \cdot [BabABababA, aaBAAb] = a(baBABAbaBabA)^3 A$$

expresses a conjugate of w^3 as a product of three commutators, and defines an extremal surface virtually bounding w. Consequently $\text{scl}(w) = 5/6$. On the other hand, it turns out that elements in free groups with half-integral scl are very *common*; see § 4.1.9 and § 4.2.

The algorithm as described above is hopelessly inefficient for all but a handful of words. In the next section we will describe a much more dramatic improvement, resulting in a polynomial time algorithm.

4.1.7. A polynomial time algorithm to calculate scl **in free groups.** An extremal surface is built from rectangles and polygons. The number of rectangle types is quadratic in the length of w, but the number of polygon types is usually of the order $|w|!$ so a naive implementation of the algorithm described in § 4.1.6 is useless for words of length 20 or more. The problem is the explosion of combinatorial types of polygons with large numbers of sides.

A polygon with many sides is the combinatorial analogue of a critical point of high index — a region in a surface with a high concentration of negative curvature. The basic idea is that a polygon with more than 4 branch edges can be split up, in a natural way, into polygons with 4 or fewer branch edges. For simplicity, in this section we restrict attention to alternating words in F_2, so that the cores of rectangles attached to consecutive branch edges alternate between $a^{\pm}$ or $b^{\pm}$. As an added simplification, shrink boundary edges to points, so that every (remaining) edge is a branch edge. Hence all polygons in question have an even number of sides.

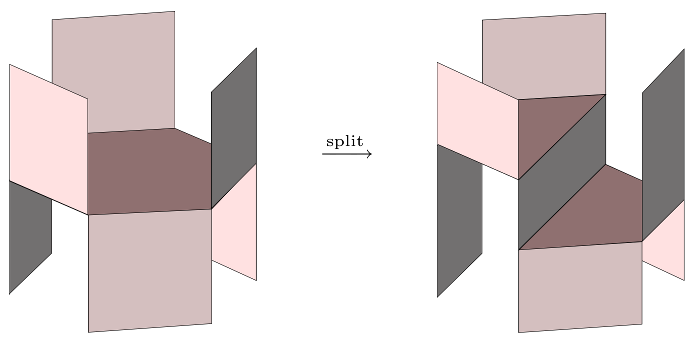


FIGURE 4.3. A hexagon can be split up into two quadrilaterals

Let P be a polygon. The (oriented) rectangles attached to P come in four kinds, depending on whether the core of the rectangle, when it moves away from P, wraps around a, A, b, B; hence if P has more than 4 sides, there are at least two pairs of rectangles of the same kind attached to P. Two (nonadjacent) rectangles of the same combinatorial kind cobound a quadrilateral Q in P. The basic idea is that the polygon P can be split up into Q and (the components of) $P - Q$; see Figure 4.3 for an example. Since the two rectangles which attach to Q wrap around the same handle of the handlebody H, we can "slide" the quadrilateral Q one third of the way around H. After a judicious sequence of slides of this kind, every remaining polygon is a quadrilateral or a bigon.

More precisely, let P be a polygon. The edges of P are labeled by a, A, b, B. All but at most one of the a edges can be paired up, resulting in a union of pairwise disjoint a-quadrilaterals $Q_a \subset P$. Do this pairing in such a way that each region of $P - Q_a$ has at most two boundary edges in ∂Q_a, or else one boundary edge in ∂Q_a and at most one unpaired a edge, then slide the Q_a quadrilaterals 1/3 of the way around the a handle. For each component P' of $P - Q_a$, pair up all but one of the A edges, resulting in a union of pairwise disjoint A-quadrilaterals $Q'_A \subset P'$. Do this pairing in such a way that each region of $P' - Q'_A$ has at most two boundary edges in $\partial Q'_A$, or else one boundary edge in $\partial Q'_A$ and at most one unpaired A edge, then slide the Q'_A quadrilaterals 1/3 of the way around the A handle.

By construction, each component P'' of $P' - Q'_A$ has at most 8 edges, half of which are b or B edges. If P'' has 4 or 2 edges, we leave it alone. If it has 6 edges, there are (without loss of generality) at least 2 b edges which span a quadrilateral Q''_b. In this case, slide the Q''_b quadrilateral 1/3 of the way around the b handle and observe that $P'' - Q''_b$ is the union of a quadrilateral and a bigon. Otherwise, suppose P'' has 8 edges. Suppose there are a pair of antipodal b or B edges. Then these span a quadrilateral Q''_b or Q''_B, and the complement in P'' is a union of two quadrilaterals. Otherwise, there are a pair of adjacent b edges and a pair of adjacent B edges spanning disjoint quadrilaterals Q''_b and Q''_B so that $P'' - Q''_b - Q''_B$ is a single quadrilateral and two bigons. In every case, after sliding Q''_b and Q''_B quadrilaterals 1/3 of the way around the b and B handles, we have achieved the desired reduction.

The final result is a surface (homotopic to the original extremal surface) made up of quadrilaterals and bigons in E, quadrilaterals 1/3 or 2/3 of the way around the handles, and (parts of) rectangles joining them up. The number of combinatorial types of (sub-) rectangles is still quadratic in $|w|$, but now the number of polygon types is of order $O(|w|^4)$. This data can be turned into a linear programming problem in $O(|w|^4)$ variables, with $O(|w|^2)$ equations. Each equation is linear in the variables, with coefficients in the finite set $\{\pm 1, \pm 1/2, 0\}$, so the data of the problem can be encoded with $O(|w|^6)$ bits. There are several well-known polynomial time methods of exactly solving a linear programming problem. For example, Karmarkar's projective method [**122**] takes time $O(n^{3.5}L)$ to exactly solve a linear programming problem in n variables encoded in L bits.

For non-alternating words, or free groups of higher rank, one must allow a larger (but still finite) set of combinatorial polygon types; the details are very similar to the alternating case. Hence we have the following:

PROPOSITION 4.36. *Let F be a free group. There is an algorithm to compute $\mathrm{scl}(w)$ for $w \in F$ whose running time is polynomial in the word length $|w|$.*

4.1.8. Foldings. In fact, for alternating words, even more simplification is possible. The basic idea is as in the previous section. Suppose S is an extremal surface with boundary on γ which contains a polygon P with more that 4 sides (after collapsing boundary edges). Then we can split off a quadrilateral and slide it around a handle. Instead of sliding it only a third of the way, slide the quadrilateral all the way around the handle. The fact that S is π_1-injective ensures that the quadrilateral does not run into another polygon when it gets all the way around the handle. However it might easily join up with some other polygon P' along an edge, and it is not clear that the result of this quadrilateral slide has made things less complicated rather than more.

The problem can be simplified using graphs, and a procedure due to Stallings [**191**] called *folding*. We replace the map of spaces $f : S \to H$ by a map of graphs $g : \Gamma \to X$ where X is a wedge of two circles (i.e. the core of the handlebody H) and Γ is the graph with one vertex for every polygon in S and one edge for every rectangle. The map g is simplicial, taking edges to edges and vertices to vertices. Let $X' \to X$ be the two-fold covering which unwraps each handle, and $g' : \Gamma' \to X'$ the map induced by g on a suitable covering space Γ' of Γ. Note that Γ and X are homotopy equivalent to S and H respectively; since extremal maps are π_1-injective, the map g is π_1-injective, and so is g'.

The graph X' is 4-valent, with two vertices. Make X' a directed graph in the following way. At each vertex of X' there are four edges, labeled a, A, b, B. Orient the edges of X' so that at one vertex, the a, A edges are outgoing, and at the other vertex the b, B edges are outgoing.

Stallings calls a simplicial map between graphs an *immersion* when it is injective on the star of every vertex. If $p : G_1 \to G_2$ is a simplicial map between graphs which is not an immersion, Stallings shows how to modify G_1 by a sequence of moves called *folds* which do not change the image of $\pi_1(G_1)$ under p_*, so at the end of the sequence of folds the resulting map is an immersion. If p is π_1-injective, each fold is an elementary collapse: two edges of G_1 which share one endpoint in common, and map to the same edge of G_2, are identified. The result of a maximal sequence of folds is well-defined independent of the choice of the sequence of folds. In fact, let $\tilde{G}_2$ denote the universal cover of G_2, which is a tree. Then $p_*(\pi_1(G_1))$ acts on $\tilde{G}_2$, and there is a unique minimal invariant subtree, whose quotient is isomorphic to the maximal folding of G_1.

In our case, since the map $g' : \Gamma' \to X'$ is already π_1-injective, each fold is an elementary collapse. There are two kinds of folds, distinguished by the orientation on X': graphically, we can perform a fold when a $\vee$ subgraph of Γ' maps to a single edge of X', by identifying the two edges of the $\vee$. If the vertex of the $\vee$ maps to the initial vertex of the directed edge of X', we say this is a *positive fold*, otherwise a *negative fold*; by abuse of notation, we say that a $\vee$ admits a positive fold, and a $\wedge$ admits a negative fold. Since g' is π_1-injective, a $\vee$ and a $\wedge$ can share at most one edge in common, and therefore consecutive positive and negative folds can be performed in either order. Hence we can arrange to perform all positive folds first, then all negative folds, in some maximal sequence of folds.

Let $f' : S' \to H'$ be the associated maps of double covers. Note that the composition $S' \to H' \to H$ is extremal if S is. The orientation on X' gives an unambiguous sense to what it means to slide a quadrilateral of S' over a handle of H' in the positive direction. If S' has a polygon P with at least 6 edges, then we can slide some sub-quadrilateral Q of P in the positive direction. The effect of this on the graph X' is to perform a positive fold and then the inverse of a negative fold. In other words, after sliding finitely many quadrilaterals of S', we can arrange matters so that the graph Γ' admits a maximal folding sequence with no positive folds. But such a graph admits no positive folds at all, and therefore Γ' represents a surface S' in which no polygon has more than 4 edges. In words: if w is an alternating word in F_2, some extremal monotone surface for w contains no polygons with more than 4 branch edges.

In the case that the rank is bigger than 2, replace H by a union of genus 1 solid handlebodies glued along their splitting disks as in § 4.1.4. The associated graph X is a wedge of n circles, and X' is a $2n$-valent directed graph with two vertices, at each of which there are n incoming edges and n outgoing edges. If S as above contains a polygon P with at least $2n+2$ edges, we can slide a sub-quadrilateral in the positive direction, thus performing a positive fold and the inverse of a negative fold on Γ. After sliding all quadrilaterals as far as they will go in the positive direction, the resulting graph Γ' admits no positive folds, and therefore the surface S' contains no polygons with more than $2n$ branch edges.

Hence we have proved:

Proposition 4.37. *Let w be an alternating word in F_n. Then some extremal surface for w contains no polygons with more than $2n$ branch edges.*

This proposition leads to a further dramatic reduction in the time needed to compute scl on alternating words, especially in F_2. The resulting algorithm has been implemented in the program `scallop`, whose source is available from [**39**]. In practice the runtime is quite modest, taking on average about 6 seconds on a late 2008 MacBook Pro to compute scl on an alternating word of length 60 in F_2.

Remark 4.38. A decomposition of a surface into rectangles and polygons determines a vector field on the surface with a saddle singularity for every 4-gon, and an n-prong monkey saddle singularity for every $2n+2$-gon. Such data defines a branched Euclidean metric on the surface where the negative curvature is concentrated at the singularities. Bounding the number of sides of the polygons is the combinatorial equivalent of finding two sided curvature bounds for a smooth surface. A closed least area surface in a non-positively curved 3-manifold has two sided curvature bounds, but for a surface with boundary, there are no such *a priori* lower bounds. Thus it is perhaps somewhat surprising that such uniform lower bounds on the complexity of the polygons in an extremal surface can be obtained, independent even of γ.

4.1.9. Gaps, limits, tongues. An alternating word in F_2 has length $4n$ for some n. There are $2\cdot(2n!)^2/(n!)^4$ alternating words of length $4n$, but after applying conjugation and anti-involutions $a \leftrightarrow A$ and $b \leftrightarrow B$ if necessary, we may assume the word starts with ab.

Computer experiments using `scallop` reveal unexpected structure in the scl spectrum of F_2.

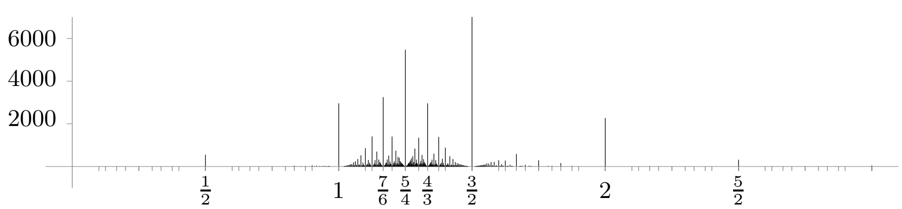


Figure 4.4. Values of scl on 50,000 random alternating words of length 36. The horizontal axis is scl and the vertical axis is frequency (the spike at 3/2 is attenuated to fit in the figure).

Figure 4.4 is a histogram of values of scl on random alternating words of length 36. There are several conspicuous features of this plot, including:

(1) the existence of a spectral gap between 0 and 1/2 (discussed in § 4.3.4)
(2) the indiscreteness of the set of values attained
(3) the relative abundance of elements whose scl has a small denominator

The self-similarity of the histogram suggests the existence of a *power law* for the frequency of elements with scl a given rational, of the form $\text{freq}(p/q) \sim q^{-\delta}$ where $\delta \sim 2$ in this example. This self-similarity persists on a fine scale (see Figure 4.5). Co-ordinates of the spikes are obtained by Farey addition of nearest spikes, after multiplying numerators by 2.

Similar power laws occur in dynamical systems, e.g. in the phenomenon of "frequency locking" for coupled nonlinear oscillators. One of the best-known examples

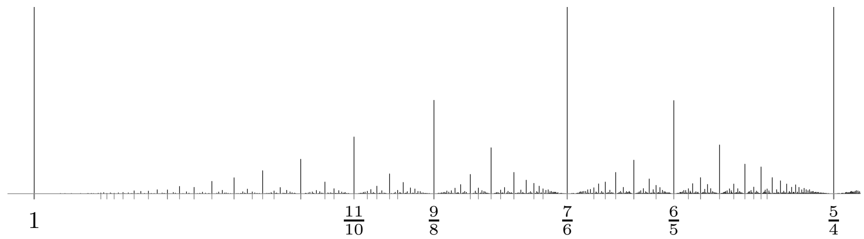


FIGURE 4.5. A stretched scaled excerpt from Figure 4.4.

is that of *Arnold tongues* (see [**2**]). For $K \in [0,1]$ and $\omega \in S^1$ define a function $f_{K,\omega} : S^1 \to S^1$ by

$$f_{K,\omega}(\theta) = \theta + \omega - K \sin 2\pi\theta$$

This is a homeomorphism for $K \leq 1$, and one can look at the rotation number $\text{rot}(f_{K,\cdot})$ as a function of ω for varying K. In fact, for $K > 0$, the set of ω for which this rotation number is a given rational is a nonempty closed interval, and these intervals expand as $K \to 1$ to completely fill out the circle (in measure). Following [**116**] we define $\Delta(p/q)$ to be the length of the interval of values ω for which the rotation number is equal to p/q for $K = 1$. Jensen et. al. [**116**] found experimentally that the heights $\Delta(p/q)$ obey a power law, with $\Delta(p/q) \sim q^{-\delta}$ for $\delta = 2.292 \pm 3.4 \times 10^{-3}$.

The indiscreteness of the spectrum is more evident when one includes non-alternating words.

EXAMPLE 4.39. For positive integers n, m define $s(n,m) = \text{scl}([a,b^n][a,b^{-m}])$. Then $s(n,m) = s(m,n)$, and $s(n,m) = 1 - 1/t(n,m)$ where $t(n,m) = t(n/d, m/d)$ if $\gcd(n,m) = d$, and

$$t(n,m) = \max(2n - 2m, n) \text{ if } \gcd(n,m) = 1 \text{ and } n > m$$

In particular, every value of $\mathbb{Q}$ mod $\mathbb{Z}$ is achieved in scl of F_2 (and therefore in any nonabelian free group).

For a proof and a (partial) explanation, see [**46**]. On the other hand, not every positive rational number occurs as a value of scl in a free group. As has been remarked before, $\text{scl}(w) \geq 1/2$ for all nontrivial $w \in [F_2, F_2]$, and the value of $1/2$ is realized on every commutator. Experimentally, there appears to be another gap in the spectrum between $1/2$ and $7/12$, then a gap between $7/12$ and $5/8$, with the first accumulation point of the set $\text{scl}([F_2, F_2])$ at $3/4$ (of course, each nonzero value is achieved on infinitely many conjugacy classes; compare with Theorem 3.11). Finally, experiments suggest that *every* rational number ≥ 1 is in the scl spectrum.

4.1.10. Injective, extremal, isometric maps. A map $f : \pi_1(S) \to G$ of a surface group into a group G is *injective* if it is a monomorphism, and *extremal* if it realizes the infimum of $-\chi^-(S)/2n(S)$ for its boundary. Say it is *isometric* if $\text{scl}(f(a)) = \text{scl}(a)$ for all $a \in [\pi_1(S), \pi_1(S)]$ (note that injective and isometric maps make sense between arbitrary groups). There are inclusions

$$\text{isometric} \subset \text{extremal} \subset \text{injective}$$

It is an interesting problem to delineate precisely the difference between these three natural classes of surfaces.

EXAMPLE 4.40. Any automorphism is isometric.

EXAMPLE 4.41. If an inclusion $f : G \to H$ splits, then f is isometric.

EXAMPLE 4.42. For any nonzero integers n, m the map $F_2 \to F_2$ sending $a \to a^n$ and $b \to b^m$ is isometric (see [**46**] for a proof).

EXAMPLE 4.43 (once punctured torus). Any map $f : F_2 \to F_2$ has image which is either cyclic or injective. Furthermore, since $1/2$ is a lower bound on nontrivial elements for scl in a free group, every injective map from F_2 to itself (or to any free group) is extremal.

EXAMPLE 4.44 (high distance Heegaard splittings). The following example was inspired by an idea of Geoff Mess. A *Heegaard splitting* exhibits a closed 3-manifold M as a union of two handlebodies H_1, H_2 glued along a surface S. Recall (Definition 3.69) the definition of the complex of curves $\mathcal{C}(S)$. Each handlebody H_i determines a subcomplex $\mathcal{C}(H_i)$ in the complex of curves $\mathcal{C}(S)$ consisting of isotopy classes of essential simple closed curves in S which bound disks in H_i. The *distance* of a Heegaard splitting is the length of the shortest path in the 1-skeleton of $\mathcal{C}(S)$ from a vertex in $\mathcal{C}(H_1)$ to a vertex in $\mathcal{C}(H_2)$. 3-manifolds with Heegaard splittings of arbitrarily high distance and genus exist, and are easy to construct (see e.g. Hempel [**108**]). Let M be a 3-manifold with a Heegaard splitting of genus at least 3 and distance at least 2. Let $\alpha \subset S$ bound a disk in H_1, and separate S into two subsurfaces of different genus. Since the distance of the splitting is at least 2, every simple essential loop in S which bounds a disk in H_2 must intersect α non-trivially. Hence, by the loop theorem (see [**107**] p. 39) the components of $S - \alpha$ are π_1-injective in H_2. Since H_2 is a handlebody, $\pi_1(H_2)$ is free (of rank ≥ 3). This example shows that there are (many) injective surfaces in free groups which are not extremal. Note that a free group of any rank can be included into a free group of rank 2, so there are examples of injective, non-extremal surfaces in free groups of any rank.

EXAMPLE 4.45. Another example is due to Justin Malestein, based on Witt identities. Let F be a free group with generators $x_1, x_2, \cdots, x_n$ for some large n. Define

$$s_i = \begin{cases} x_1 & \text{if } i = 1, 2 \\ x_{2+(i-1)/2} x_{1+(i-1)/2} x_{2+(i-1)/2}^{-1} & \text{if } i > 2 \text{ is odd} \\ x_{1+i/2}[[\cdots[x_1, x_2], x_3], \cdots, x_{-1+i/2}] x_{1+i/2}^{-1} & \text{if } i > 2 \text{ is even} \end{cases}$$

Then one can verify that for each $g \leq n/2$, the elements $s_1, s_2, \cdots, s_{2g}$ generate a free subgroup of F of rank $2g$, and moreover that there is an identity

$$[s_1, s_2] \cdots [s_{2g-1}, s_{2g}] = [[x_1, \cdots, [x_{g-1}, x_g] \cdots], x_{g+1}]$$

thus exhibiting a genus g surface group and a genus 1 surface subgroup of F with the same boundary.

For example, if $g = 2$, one has the identity

$$[s_1, s_2][s_3, s_4] = [x_1, x_2] x_3 [x_2, x_1] x_3^{-1} = [[x_1, x_2], x_3]$$

Since every subgroup of a free group is free, there are no injective maps from *closed* surface groups to free groups. However, we can use extremal surfaces to construct injective maps from closed surface groups to many groups obtained from free groups by simple procedures.

A well-known question due to Gromov [**98**] is the following:

QUESTION 4.46 (Gromov). *Does every* 1*-ended word-hyperbolic group contain a closed hyperbolic surface subgroup?*

This question seems to be far beyond the reach of current technology. Nevertheless, as an application of the Rationality Theorem, we can find such surfaces in certain groups, obtained as graphs of free groups amalgamated along cyclic subgroups (for an introduction to the theory of graphs of groups, see e.g. Serre [**187**], especially Chapter 1).

THEOREM 4.47. *Let G be a finite graph of free groups, amalgamated along cyclic subgroups.*

1. *Every $\alpha \in H_2(G;\mathbb{Z})$ has a multiple which is represented by a π_1-injective map of a closed surface (which may be disconnected).*
2. *The unit ball of the Gromov (pseudo-)norm on $H_2(G;\mathbb{R})$ is a finite sided rational polyhedron.*
3. *Let $g_1, g_2, \cdots, g_n \in G$ be conjugate into (free) vertex subgroups of G. Then* scl *is piecewise rational linear on $\langle g_1, \cdots, g_n\rangle \cap B_1^H(G)$, and every rational chain in this subspace rationally bounds an extremal surface.*

REMARK 4.48. If some homology class in G is represented by a $\mathbb{Z}\oplus\mathbb{Z}$, the Gromov pseudo-norm on $H_2(G;\mathbb{R})$ is degenerate. In this case, the proposition should be construed as saying that $\|\cdot\|_1$ is a non-negative convex piecewise rational linear function. On the other hand, if G is word-hyperbolic, $\|\cdot\|_1$ is a genuine (polyhedral) norm.

REMARK 4.49. In contrast with the case of a 3-manifold, the norm $\|\cdot\|_1$ does not generally take integral values on $H_2(G;\mathbb{Z})$.

We give the sketch of a proof; for details, see [**43**].

PROOF. Since G is a graph of free groups amalgamated along cyclic subgroups, there is a $K(G,1)$, denoted X, obtained as a union $X = H \cup A$, where H is a disjoint union of handlebodies, and A is a disjoint union of annuli attached along their boundary to essential loops in H (in fact, this can be taken to be the definition of a graph of free groups amalgamated over cyclic subgroups). If H_i is a component of H, let F_i denote the corresponding (free) vertex subgroup of G. Furthermore, for each i, let $\partial_i A$ denote the components of ∂A attached to H_i. We think of each $\partial_i A$ either as a set of free homotopy classes of loops in H_i, or as a set of conjugacy classes in F_i.

Let $\alpha \in H_1(G;\mathbb{Z})$ be given, and let $f : S \to X$ be a map of a surface representing α. After compression and a homotopy, we can insist that $f^{-1}(A)$ is a union of annuli, each of which maps to some component of A by a covering map. Write S as a union $S = T \cup U$, where $U = f^{-1}(A)$, and $T = \cup_i T_i$ where $T_i = f^{-1}(H_i)$. The image $f_*(\partial T)$ is a chain C which can be written as a formal sum $C = \sum C_i$ where each C_i has support in H_i. By construction, $C_i \in \langle \partial_i A\rangle \cap B_1^H(F_i)$.

For each i, let T_i' be an extremal surface in H_i virtually bounding the chain C_i. By passing to common covers if necessary, we can assume that $T' = \cup T_i'$ virtually bounds C. We would like to build a surface S' by gluing up boundary components of the T_i' along covers of the cores of the annuli A. This can be accomplished by passing to a further finite cover, by Proposition 2.13. The resulting surface S' is

Gromov norm minimizing in its (projective) homology class, and is therefore π_1-injective. This proves bullet (1). Bullet (2) follows from the piecewise rational linearity of the scl norm on each $\langle \partial_i A\rangle \cap B_1^H(F_i)$.

The proof of bullet (3) is similar. Let Γ be a collection of loops representing the conjugacy classes g_i. Any admissible surface $f : S, \partial S \to X, \Gamma$ can be homotoped and compressed until $f^{-1}(A)$ is a union of annuli, each of which maps to some component of A by a covering map. Then the claim follows as above by the fact that scl is piecewise rational linear on each subspace of the form $\langle \partial_i A \cup (\Gamma \cap H_i)\rangle \cap B_1^H(F_i)$. $\square$

Example 4.50. Let F be a free group and Z a nontrivial cyclic subgroup, contained in $[F, F]$. Let G be obtained from two copies of F by amalgamating them along Z; i.e. $G = F *_Z F$. Topologically, if γ is a loop in H representing the conjugacy class of a generator of Z, the group G is the fundamental group of the space X obtained by gluing two copies of H together along γ. There is an involution ι on X which exchanges the two copies of H, and fixes γ. If $f : S \to H$ is an extremal surface which (rationally) bounds some cover of γ, there is a map Df from the double DS to X obtained by reflecting f across γ using ι. By construction, the map is injective, and realizes the Gromov norm on some multiple of the generator of $H_2(G; \mathbb{Z})$.

Example 4.51. Let S be a closed orientable surface, and let $A \subset S$ be an essential annulus in S. Let $g_1, \cdots, g_n \in \pi_1(S)$ be conjugacy classes represented by loops γ_i in $S - A$. Then scl is piecewise rational linear on the subspace $\langle g_1, \cdots, g_n\rangle \cap B_1^H(\pi_1(S))$.

Example 4.52. Let G be a graph of free groups amalgamated along cyclic subgroups. Then every finite index subgroup G' is also a graph of free groups amalgamated along cyclic subgroups. So if some finite index G' as above has nontrivial H_2, it contains a closed surface subgroup, and therefore so does G. Cameron Gordon and Henry Wilton [**94**] have several interesting criteria to guarantee this condition.

Remark 4.53. Compare the proof of Theorem 4.47 with the proofs of Theorem 2.93 and Theorem 2.101.

4.2. Geodesics on surfaces

The results of § 4.1 let us compute scl and construct extremal surfaces for arbitrary elements and chains in $B_1^H(F)$ where F is a free group. Bavard duality implies the existence of extremal quasimorphisms with rational values and rational defects, but such quasimorphisms are apparently quite elusive, and it remains a challenging problem to try to construct them. The most constrained extremal quasimorphisms (and therefore the easiest to find) should be those dual to top dimensional faces of the scl polyhedron; but for an infinite dimensional polyhedron, it becomes complicated even to give a precise definition of a top dimensional face.

However, it turns out that there *are* some naturally occurring top dimensional faces of the scl polyhedron for F a free group. More precisely, for each realization of F as $\pi_1(S)$ where S is an oriented surface (necessarily of negative Euler characteristic), there is a top dimensional face π_S of the scl norm ball. Moreover, the projective class of the chain ∂S in $B_1^H(F)$ intersects this face in its interior, and the unique homogeneous quasimorphism dual to this face (up to scale and elements of $H^1(F)$), is the *rotation quasimorphism* associated to the natural action of $\pi_1(S)$ on the circle at infinity of hyperbolic space coming from any choice of hyperbolic structure on S. This is Theorem 4.78, to be proved in the sequel.

4.2.1. Self-intersections. Fix the following conventions. Let S be an orientable surface of finite type (usually compact and connected with nonempty geodesic boundary) with $\chi(S) < 0$. If we fix a hyperbolic structure on S, then every free homotopy class of loop has a unique (unparameterized) geodesic representative. If $a \in \pi_1(S)$, and $[a]$ denotes the conjugacy class of a, then we let $\gamma(a)$ denote the geodesic in the free homotopy class determined by $[a]$. If we want to refer specifically to the hyperbolic metric g on S, we write $\gamma(a, g)$.

We recall from § 3.5.3 the notation $\text{cr}(a)$ for the number of self-intersections of $\gamma(a)$ in S (i.e. the *crossing number*). Our discussion in § 3.5.3 was brief and somewhat sketchy; we are more careful now.

The combinatorics of the geodesic $\gamma(a)$ in S does not depend on the choice of hyperbolic structure when $\text{cr}(a) \leq 2$. But when γ has 3 or more self-intersections the combinatorics of γ may (and usually will) depend on the geometry of S. In particular, three local sheets might undergo a "Reidemeister 3" move; see Figure 4.6.

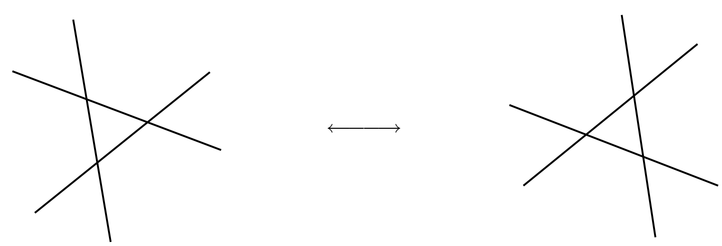


FIGURE 4.6. A Reidemeister 3 move

More subtly, a geodesic representative might not be in general position, and a "coincidental" triple point might be stable under deformations of the hyperbolic structure.

EXAMPLE 4.54 (Hass–Scott [**104**]). This example is a straightforward variation on Example 5 from [**104**]. A hyperbolic once-punctured torus T has an isometric involution which fixes the boundary and three interior (Weierstrass) points. A suitable free homotopy class of loop in T invariant by this involution has a geodesic representative which is forced to go through some or all of these points, an arbitrary number of times. This example can be inserted into any non-planar hyperbolic surface.

Self-intersections and crossing number are more properly defined in terms of linking data at infinity. Let S^1_∞ denote the circle at infinity of the hyperbolic plane. Two disjoint pairs of points in S^1_∞ are said to be *linked* if each separates the other in S^1_∞. Formally, we define a self-intersection of γ as follows. Let $s : S^1 \to \gamma \subset S$ be a parameterization of γ. By abuse of notation, we say that a *lift* of s is a map $\tilde{s} : \mathbb{R} \to \mathbb{H}^2$ which intertwines the covering projections $\mathbb{R} \to S^1$ and $\mathbb{H}^2 \to S$. A *self-intersection* of γ is an unordered pair of lifts $\tilde{s}_1, \tilde{s}_2$ for which the endpoints of the geodesics $\tilde{s}_1(\mathbb{R}), \tilde{s}_2(\mathbb{R})$ are linked in S^1_∞, up to the action of the deck group $\pi_1(S)$

on such pairs. Then define $\mathrm{cr}(a)$ to be the cardinality of the set of self-intersections of $\gamma(a)$.

Linking number is well-defined independent of the hyperbolic structure on S, so this notion is purely topological. For primitive geodesics in general position, the cardinality of the set of self-intersections agrees with the naive (geometric) definition of crossing number, and satisfies the desirable property $\mathrm{cr}(a^n) = n^2\mathrm{cr}(a)$.

If γ is not generic, we distinguish the abstract set of self-intersections (as defined above) from the *support* of the self-intersections, which is a finite subset of γ, and whose cardinality might depend on the hyperbolic structure on S.

4.2.2. Bounding surfaces. Assume now that S is compact, possibly with boundary. Fix a hyperbolic structure on S and an element $a \in \pi_1(S)$, and let γ denote the (oriented) geodesic corresponding to the conjugacy class of a.

For a given hyperbolic structure, γ decomposes $S - \gamma$ into a finite collection of complementary regions R_i. Each region inherits an orientation from S. Moreover, γ is decomposed by its own self-intersections into a collection of oriented segments γ_j. Finally the support of the self-intersections is a collection of oriented points v_i.

DEFINITION 4.55. Let $C_*(\gamma)$ be the chain complex (over $\mathbb{Z}$) generated by the oriented polyhedra R_i, γ_i, v_i together with the boundary components of S, with boundary maps the usual boundaries for polyhedra. Let $H_*(\gamma)$ denote the homology of this complex.

We let S_C denote the element of C_2 which is just the sum of the oriented generators of C_2, and γ_C the element of C_1 which is the sum of the oriented generators of C_1, excluding the boundary components.

Fix an open covering of S whose open sets are regular open neighborhoods U_i of the regions R_i. At least when S is closed, the Čech cohomology of the nerve of this covering (with constant coefficients) is canonically (because of orientations) Poincaré dual to C_*. In particular, there is a canonical surjective homomorphism from ordinary (Čech) homology $H_*(S;\mathbb{Z}) \to H_*(\gamma)$, and the classes $[S_C], [\gamma_C] \in H_*(\gamma)$ are the images of the corresponding elements in $H_*(S;\mathbb{Z})$. There is a similar interpretation of $H_*(\gamma)$ in Čech homology when S has boundary.

LEMMA 4.56. *The kernel of $\partial : C_2(\gamma) \to C_1(\gamma)$ is generated by S_C if S is closed, and is zero otherwise.*

PROOF. This follows by the remarks in the paragraph above, together with the fact that S is connected and orientable, and therefore $H_2(S;\mathbb{Z})$ is at most 1 dimensional, and is 0 dimensional unless S is closed. □

Since γ is closed, γ_C is a cycle. If $a \in [\pi_1(S), \pi_1(S)]$ then $[\gamma] = 0 \in H_1(S)$, so $\gamma_C = \partial A_\gamma$ for some $A_\gamma \in C_2$. If S is not closed, ∂ is injective on C_2 by Lemma 4.56, and therefore A_γ is uniquely defined. For each region R_i, let w_i denote the coefficient of the generator R_i in A_γ, so that $A_\gamma = \sum_i w_i R_i$.

Let T be a compact orientable surface, possibly with multiple boundary components, and let f be a map of pairs $f : (T, \partial T) \to (S, \gamma)$. If we put f in general position, f restricts to a proper map between open surfaces $T - f^{-1}(\gamma) \to S - \gamma$. The orientations on T and S determine a *degree*, denoted $\deg(f)$, which is an assignment of an integer to each region R_i; i.e. an element of $C_2(\gamma)$. If f is smooth, the degree of f on R_i is the signed sum of preimages of a generic point in R_i. One way of thinking of the degree is as the image of the fundamental class of the pair $(T, \partial T)$ in a suitable relative homology group.

Enumerate the components of ∂T as $\partial_i T$, and suppose that $f(\partial_i T)$ represents γ^{n_i} in $\pi_1(S)$. We define the degree of $f|_{\partial T}$ similarly, and write $\deg(\partial_i f) = n_i \gamma_C$ and $\deg(\partial f) = \sum_i n_i \gamma_C$. Write $n(T) = \sum n_i$ as above. From the definition we have

$$\partial \deg(f) = \deg(\partial f)$$

and so from Lemma 4.56, we deduce

$$\deg(f) = n(T) \cdot A_\gamma$$

providing S has nonempty boundary.

4.2.3. Area norm. Throughout this section, all surfaces under discussion are assumed to have nonempty boundary, unless we explicitly say to the contrary.

DEFINITION 4.57. For $a \in [\pi_1(S), \pi_1(S)]$ and for a fixed choice of hyperbolic metric g on S, define the *area* of $\gamma(a, g)$ by

$$\text{area}(\gamma(a,g)) = \sum_i w_i \text{area}(R_i)$$

where $A_\gamma = \sum w_i R_i$, and

$$\text{area}^+(\gamma(a,g)) = \sum_i |w_i| \text{area}(R_i)$$

If g and a are understood, we abbreviate this to $\text{area}(\gamma)$ and $\text{area}^+(\gamma)$ respectively.

From the definition there is an inequality $\text{area}^+(\gamma) \geq |\text{area}(\gamma)|$ with equality if and only if all the w_i have the same sign.

DEFINITION 4.58. If all the w_i have the same sign, then γ is *monotone*.

LEMMA 4.59. *Let a, γ be as above. Then for any hyperbolic structure g on S there is an inequality*

$$\text{scl}(a) \geq \frac{\text{area}^+(\gamma(a,g))}{4\pi}$$

PROOF. For each surface $(S_i, \partial S_i) \to (S, \gamma)$ we either compress S_i along an essential embedded loop or arc, or else we can find a pleated representative. The pleated representative defines a hyperbolic structure on S_i with totally geodesic boundary. Moreover, by definition, we have

$$\text{area}(S_i) = \sum_i \int_{R_i} \#\{f^{-1}\} d\text{area} \geq \sum_i \int_{R_i} |\deg(f) \text{ on } R_i| d\text{area} = n(S_i)\text{area}^+(\gamma)$$

By Gauss–Bonnet, $\text{area}(S_i) = -2\pi\chi(S_i)$. By Proposition 2.10, $\text{scl}(a)$ is the infimum of $-\chi(S_i)/2n(S_i)$ over all such S_i. □

Values of $\text{area}(\gamma)$ are *quantized*:

LEMMA 4.60. *For any $a \in \pi_1(S)$ and any hyperbolic metric g,*

$$\text{area}(\gamma(a,g)) \in 2\pi\mathbb{Z}$$

In particular, $\text{area}(\gamma)$ *does not depend on g.*

Proof. Let $(S', \partial S') \to (S, \gamma)$ be a pleated surface for which $n(S') = 1$. The pleated surface structure determines a decomposition of S' into an even number of ideal triangles, whose areas sum to area(S'). The Jacobian $J(f)$ is constant on each ideal triangle, and takes values in ± 1. We calculate

$$\text{area}(\gamma) = \sum_i \int_{R_i} \deg(f) \text{ on } R_i \, d\text{area} = \int_{S'} J(f) \, d\text{area}$$

which is a sum of an even number of π's and $-\pi$'s. □

In fact, the relationship between area and scl is precise enough to detect a significant amount of topological information. An immersion $f : T \to S$ between oriented surfaces is *positive* if it is orientation-preserving on each component, and *negative* if it is orientation-reversing on each component. Note that if S and T are both connected, every immersion between them is either positive or negative.

For the moment we are considering immersed loops in surfaces S. In the sequel we will consider immersed 1-manifolds. In anticipation therefore, we make the following definition.

Definition 4.61. An immersed oriented 1-manifold $\Gamma : \coprod_i S^1 \to S$ *bounds* a positive immersion $f : T \to S$ if there is a commutative diagram

$$\begin{array}{ccc} \partial T & \xrightarrow{\quad i \quad} & T \\ {\scriptstyle \partial f}\downarrow & & \downarrow{\scriptstyle f} \\ \coprod_i S^1 & \xrightarrow[\quad \Gamma \quad]{} & S \end{array}$$

for which $\partial f : \partial T \to \coprod_i S^1$ is an orientation-preserving homeomorphism. The 1-manifold Γ *virtually bounds* (or *rationally bounds*) a positive immersion as above if there is a positive integer n so that $\partial f : \partial T \to \coprod_i S^1$ is an orientation-preserving covering satisfying $\partial f_*[\partial T] = n[\coprod_i S^1]$ in homology.

The property of virtually bounding an immersed surface can be detected by stable commutator length:

Lemma 4.62. *Let $a \in \pi_1(S)$ be represented by a geodesic $\gamma \subset S$. Suppose γ virtually bounds a positive immersed surface T. Then T is extremal, and*

$$\text{scl}(a) = \text{area}(\gamma)/4\pi = -\chi(T)/2n$$

Conversely, if γ does not virtually bound a positive immersed surface, then $\text{scl}(a) > \text{area}(\gamma)/4\pi$.

Proof. Under the hypotheses of the Lemma, $n\text{area}(\gamma) = \text{area}(T)$. If γ virtually bounds a positive immersed surface T, then $\text{scl}(a) \le -\chi(T)/2n$. This gives an upper bound on scl which is equal to the lower bound in Lemma 4.59.

Conversely, let T be extremal for a (such a T exists by Theorem 4.24). If T is not homotopic to an immersion, then a pleated representative of T maps at least one ideal triangle with degree -1 and therefore $\text{scl}(a) = -\chi^-(T)/2n > \text{area}(\gamma)/4\pi$. □

Remark 4.63. By changing the orientation on γ, one sees that γ virtually bounds a negative immersed surface if and only if $\text{scl}(a) = -\text{area}(\gamma)/4\pi$.

Remark 4.64. We will see in Example 4.72 that there are examples of curves γ which do not bound an immersed surface, but have finite (disconnected) covers which *do* bound immersed surfaces.

REMARK 4.65. One direction of Lemma 4.62 is easy: an immersed surface is evidently extremal, by Bavard duality. The other direction of the proof really uses the existence of extremal surfaces, and therefore depends on Theorem 4.24.

COROLLARY 4.66. *Let $a \in \pi_1(S)$ be represented by a geodesic γ. Suppose a finite cover of γ bounds a (positive or negative) immersed surface in S. Then* $\mathrm{scl}(a) \in \frac{1}{2}\mathbb{Z}$.

PROOF. By Lemma 4.62, there is an equality $\mathrm{scl}(a) = |\mathrm{area}(\gamma)|/4\pi$. On the other hand, by Lemma 4.60, $\mathrm{area}(\gamma) \in 2\pi\mathbb{Z}$. □

REMARK 4.67. Although $\mathrm{area}(\gamma)$ does not depend on the hyperbolic metric g, the quantity $\mathrm{area}^+(\gamma(a,g))$ might. By Gauss–Bonnet, the area of a hyperbolic polygon P is

$$\mathrm{area}(P) = \pi(n-2) - \sum_i \alpha_i$$

where n is the number of vertices, and the α_i are the internal angles. Summing contributions of this kind, we see that $\mathrm{area}^+(\gamma(a,g))$ is an integral linear combination

$$\mathrm{area}^+(\gamma(a,g)) = \sum_p n(p,g)\alpha(p,g) + \text{topological term}$$

where the topological term is in $\pi\mathbb{Z}$, where the sum is taken over points p at which γ crosses itself, where $\alpha(p,g)$ is the angle γ makes with itself at p, and where each $n(p,q)$ is an integer. The $n(p,q)$ are not constant, since they might change sign under a deformation in which some (necessarily simply-connected) region becomes degenerate and changes orientation.

It would be interesting to study $\mathrm{area}^+(\gamma(a,\cdot))$ for each $a \in \pi_1(S)$ as a function on Teichmüller space, and to characterize its range algebraically.

4.2.4. Area and rotation number. We give a reinterpretation of $\mathrm{area}(\gamma)$ in terms of *rotation numbers* which gives another explanation of the quantization of area proved in Lemma 4.60.

A hyperbolic structure and an orientation on S determines a representation $\rho : \pi_1(S) \to \mathrm{PSL}(2,\mathbb{R})$ which is unique up to conjugacy. There is a universal central extension

$$0 \to \mathbb{Z} \to \widetilde{\mathrm{SL}}(2,\mathbb{R}) \to \mathrm{PSL}(2,\mathbb{R}) \to 0$$

with extension class $[e] \in H^2(\mathrm{PSL}(2,\mathbb{R});\mathbb{Z})$.

If G is any group, and $\rho : G \to \mathrm{PSL}(2,\mathbb{R})$ is a representation, $[e]$ pulls back by ρ^* to define an element $\rho^*([e])$ of $H^2(G;\mathbb{Z})$. If ρ is understood, we abbreviate this by $[e]$ where no confusion can arise. There is an elegant description of e at the level of chains, due to Thurston [**197**]. The group $\mathrm{PSL}(2,\mathbb{R})$ acts on S^1_∞ by orientation-preserving homeomorphisms. Let $p \in S^1_\infty$ be arbitrary. If $g_1, g_2 \in G$ then define

$$e(g_1,g_2) = \begin{cases} \frac{1}{2} & \text{if } p, g_1(p), g_2(p) \text{ is positively ordered} \\ -\frac{1}{2} & \text{if } p, g_1(p), g_2(p) \text{ is negatively ordered} \\ 0 & \text{if } p, g_1(p), g_2(p) \text{ is degenerate} \end{cases}$$

More geometrically, e is $\frac{1}{2\pi}$ times the (signed) hyperbolic area of the ideal triangle spanned by $p, g_1(p), g_2(p)$. Note that e is a bounded 2-cocycle, with norm $1/2$. If $f : (S', \partial S') \to (S, \gamma)$ is a pleated surface with $n(S') = 1$, then $f_*(\partial S')$ fixes points in S^1_∞, and therefore there is a well-defined relative cocycle f^*e whose evaluation $f^*e([S'])$ is $\frac{1}{2\pi}$ times the signed sum of areas of the ideal triangles of S'; i.e. $f^*e([S']) = \mathrm{area}(\gamma)/2\pi$.

If $\rho^* e$ is trivial in $H^2(G;\mathbb{Z})$ then ρ lifts to $\tilde{\rho} : G \to \widetilde{\mathrm{SL}}(2,\mathbb{R})$. As in § 2.3.3 there is a well-defined homogeneous quasimorphism rot on G determined by the choice of a lift $\tilde{\rho}$. Different lifts are parameterized by choices of $H^1(G)$. In particular, $\tilde{\rho}$ is well-defined on $[G,G]$. As bounded cohomology classes, $-[\delta \text{ rot}] = [e]$ in $H^2_b(G;\mathbb{R})$. Here the minus sign appears because of the negative curvature of a hyperbolic surface. In fact, for any closed hyperbolic surface T, there is an equality $e([T]) = -\chi(T)$.

LEMMA 4.68. *With definitions as above, for each a in the commutator subgroup there is an equality*

$$\text{area}(\gamma(a)) = -2\pi \, \text{rot}(a)$$

PROOF. Let $f : (S', \partial S') \to (S, \gamma)$ be a pleated surface with $n(S') = 1$. Then

$$\text{area}(\gamma(a))/2\pi = e(f_*[S']) = -(\delta \text{rot})(f_*[S']) = -\text{rot}(f_*[\partial S']) = -\text{rot}(\gamma)$$

□

Since S is a complete hyperbolic surface, every element is either hyperbolic or parabolic, and therefore has a fixed point in S^1_∞. This implies that rot takes on only integral values. This explains the quantization observed earlier.

REMARK 4.69. Lemma 2.58 says that for any homogeneous quasimorphism ϕ, there is an inequality $D(\phi) \le 2\|[\delta\phi]\|_\infty$. The discussion above shows that this inequality is an equality when ϕ is the rotation quasimorphism associated to a hyperbolic structure on a noncompact surface.

In fact, for any group G and any representation $\rho : G \to \text{Homeo}^+(S^1)$, we can pull back the Euler class to obtain $[e_\rho] \in H^2_b(G;\mathbb{R})$. After passing to a central extension if necessary, we can assume $[e_\rho]$ is trivial in ordinary H^2, and obtain a rotation quasimorphism rot_ρ with $[\delta \text{rot}_\rho] = [e_\rho]$.

PROPOSITION 4.70. *With notation as above, there is an equality $D(\text{rot}_\rho) = 2\|[e_\rho]\|_\infty$.*

PROOF. We give the sketch of a proof. If G has a finite orbit, then it preserves an invariant probability measure concentrated on this orbit, and therefore rot_ρ is a homomorphism, and $[e_\rho]$ is trivial in $H^2_b(G;\mathbb{R})$. Otherwise, the action is semi-conjugate to a minimal action (i.e. one in which every orbit is dense). A minimal action is either conjugate to an action by rotations (in which case rot_ρ is a homomorphism) or has a finite cyclic centralizer. Quotienting S^1 by the action of the centralizer produces a new minimal action, and multiplies both $[e_\rho]$ and rot_ρ by the same number.

So assume the action is minimal with trivial centralizer. The Milnor–Wood inequality gives $\|[e_\rho]\|_\infty \le 1/2$ for any action. On the other hand, such an action has the following *compressibility* property: for any closed interval $I \subset S^1$ and any nonempty open set $U \subset S^1$, there is $g \in G$ for which $g(I) \subset U$; a proof of this fact (and the nontrivial assertions in the previous paragraph) follows from Thurston [**197**], Theorem 2.7. Choose disjoint nonempty connected open sets U_1, U_2, V_1, V_2 for which a pair of points in U_1 and V_1 link a pair of points in U_2 and V_2. Let g take $S^1 - V_1$ into U_1, and let h take $S^1 - V_2$ into U_2. Then the action of $\langle g, h \rangle$ is semi-conjugate to an action arising from a hyperbolic structure on a once-punctured torus. Consequently $\text{rot}_\rho([g,h]) = 1$ and therefore $D(\text{rot}_\rho) \ge 1$.

Hence

$$1 \ge 2\|[e_\rho]\|_\infty \ge D(\text{rot}_\rho) \ge 1$$

and the Proposition is proved. □

Note that the method of proof shows that any group acting on a circle either preserves a probability measure, or contains a nonabelian free subgroup. In the literature this fact is frequently attributed to Margulis [**145**], who seems not to have been aware of the work of Thurston and others.

Lemma 4.62 and Lemma 4.68 taken together show that an element a in the commutator subgroup of $\pi_1(S)$ is represented by a geodesic which virtually bounds an immersed surface in S if and only if rot is an extremal quasimorphism for a. It is convenient to extend this observation to rational chains in B_1^H.

Let $F = \pi_1(S)$, and let $C = \sum t_i a_i$ be a chain in $B_1^H(F)$. Each a_i is represented by a geodesic γ_i in S, so the chain C is represented by a "weighted" union Γ of geodesics in S. The support of Γ decomposes S into regions R_i. For each region R_i, choose an arc α_i from ∂S to R_i, and look at the (weighted) algebraic intersection $\alpha_i \cap \Gamma$. The condition that C is homologically trivial implies that this algebraic intersection number is independent of the choices involved. In the special case that C consists of a single element a, this intersection number is equal to the weight w_i as defined in Definition 4.57. Then define

$$\text{area}(\Gamma) = \sum_i (\alpha_i \cap \Gamma)\text{area}(R_i)$$

Then one has the analogue of Lemma 4.68, namely $\text{area}(\Gamma) = -2\pi \sum t_i \text{rot}(a_i)$. If the coefficients of C are rational, then after multiplying through by a large integer we can assume that the coefficients are integers, and we can think of Γ as a signed sum of simple geodesics. Lemma 4.62 holds for such Γ, and with the same proof; i.e. a weighted union of geodesics Γ representing a chain C virtually bounds a positive immersed surface if and only if $\text{area}(\Gamma) = 4\pi\,\text{scl}(C)$. Putting these two facts together gives the following proposition:

PROPOSITION 4.71. *Let S be an oriented surface with boundary. Let C be a rational chain in $B_1^H(F)$ represented by a weighted sum of geodesics Γ. Then Γ virtually bounds a (positive or negative) immersed surface in S if and only if rot_S is an extremal quasimorphism for C; i.e. if and only if $\text{scl}(C) = |\text{rot}_S(C)|/2$.*

EXAMPLE 4.72. "Virtually bounds" in Proposition 4.71 cannot in general be improved to "bounds". Consider the immersed curve $\gamma \subset S$ in Figure 4.7, where S is a once-punctured surface of genus 2. The curve γ can be realized by a geodesic in any hyperbolic structure on S.

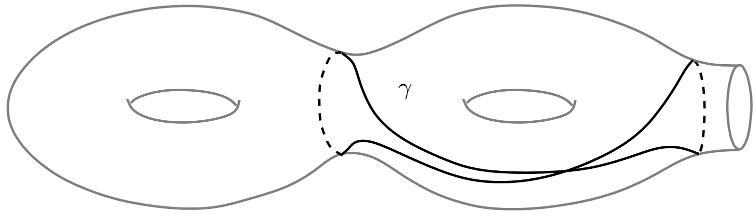


FIGURE 4.7. The loop γ does not bound an immersed surface, but two copies of γ do

The disconnected cover consisting of two copies of γ bounds an immersed surface of genus 4 with two boundary components, which each wrap once around γ. By Lemma 4.62 there is an equality $\text{scl}([a_1, b_1]^2[a_2, b_2]) = 2$ in F_4 (note that this also follows as a special case of the (free) product formula, i.e. Theorem 2.93). Since the value of scl is not of the form $1/2 + \text{integer}$, γ does not bound an immersed surface.

4.2.5. Rotation number and counting quasimorphisms. In this section, let $S_{1,1}$ denote a once-punctured torus, so that $\pi_1(S_{1,1}) = F_2$, with standard generators a, b. The function $\text{rot}_{1,1} : F_2 \to \mathbb{Z}$ is defined as above, with respect to some complete hyperbolic structure on $S_{1,1}$, and some choice of lift on the generators. Since different lifts agree on the commutator subgroup, the function $\text{rot}_{1,1}$ is well-defined in Q/H^1. One way to fix a lift is to insist that the lifts of a and b fix points, and therefore satisfy $\text{rot}_{1,1}(a) = \text{rot}_{1,1}(b) = 0$. We follow this convention in the sequel.

It turns out that we can give a simple formula for $\text{rot}_{1,1}$ in terms of the Brooks counting quasimorphisms (see § 2.3.2). Recall that for each string σ, the function H_σ counts the number of copies of σ minus the number of copies of σ^{-1}, and $\overline{H}_\sigma$ denotes its homogenization.

REMARK 4.73. In fact, in this section we only consider strings σ of length 2 with distinct letters. For such strings, the "little" and the "big" counting functions and their associated quasimorphisms h_σ and H_σ are equal.

LEMMA 4.74.

$$\text{rot}_{1,1} = \frac{1}{4}\left(\overline{H}_{ab} + \overline{H}_{ba^{-1}} + \overline{H}_{a^{-1}b^{-1}} + \overline{H}_{b^{-1}a}\right)$$

PROOF. The proof is a modification of Klein's ping-pong argument, lifted from the circle to the line. The disk D can be decomposed into 5 regions, one of which, P, is an ideal square which is a fundamental domain for F_2, and the other 4 are neighborhoods of the attracting fixed points of the elements a, b, a^{-1}, b^{-1} respectively. Call these neighborhoods N_a, N_b, N_A, N_B. Given a reduced word $\sigma \in F_2$, and a point $p \in P$, the image $\sigma(p) \in N_w$ where w is the last letter of σ. We can glue $\mathbb{Z}$ copies of each of the regions N_a etc. onto $\mathbb{R}$ in such a way that the union of $\mathbb{R}$ with these regions is the universal cover of $D - P$. Denote this union by E. See Figure 4.8.

These lifted neighborhood regions break up $\mathbb{R}$ into "units", with four units to each lift of a fundamental domain for S^1. We can lift the itinerary of p (except for p itself) under the subwords of σ to an itinerary in E. One sees that every time the letter b appears in σ, the itinerary moves up one unit if the preceding letter was a, and down one unit if the preceding letter was a^{-1}, and similarly for other allowable 2-letter combinations. The rotation number is 1/4 the number of units, proving the formula. □

This has a particularly simple and interesting interpretation in terms of the graphical calculus introduced in § 2.2.4. A cyclically reduced element in $[F_2, F_2]$ determines a loop in the square lattice without backtracking. Such a loop may be "smoothed" at the corners to determine an immersed curve in the plane. Every anticlockwise turn contributes $1/4$ to $\text{rot}_{1,1}$, whereas every clockwise turn contributes $-1/4$. Hence $\text{rot}_{1,1}$ is just the *winding number* of the immersed curve associated to an element.

The following corollary illustrates the power of this technique.

COROLLARY 4.75. *Let $g \in F_2$ be a commutator, and let γ_g be the geodesic representative of the conjugacy class of g in T, a hyperbolic once-punctured torus.*

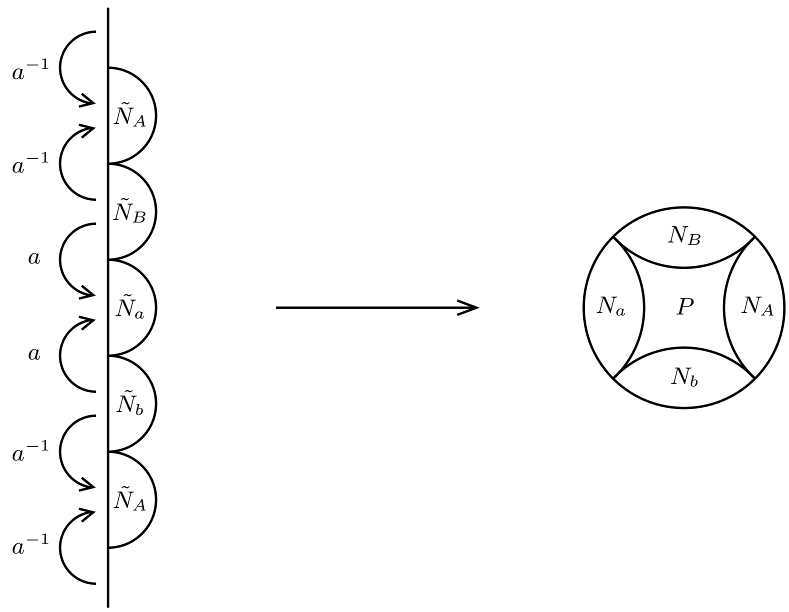


FIGURE 4.8. a moves points in $\tilde{N}_b$ regions up one unit, and a^{-1} moves such points down one unit. Furthermore, a moves points in $\tilde{N}_B$ regions down one unit, and a^{-1} moves such points up one unit. A similar relation holds with a and b interchanged.

Let w_g be the loop in the square lattice in $\mathbb{R}^2$ corresponding to the (cyclically) reduced representative of g. Then γ_g bounds an immersed surface in T if and only if the winding number of w_g is ± 1.

PROOF. Since g is a commutator, there is a map $f : T \to T$ taking the boundary to γ_g. Replace f by a pleated representative. The (algebraic) area of $f(T)$ is $-2\pi\ \mathrm{rot}_{1,1}(g) = -2\pi\ \mathrm{wind}(w_g)$, so if the winding number is ± 1, this pleated representative is an immersion. Conversely, if $\mathrm{wind}(w_g) = 0$, the algebraic area is zero, so no map f as above can be an immersion. □

A similar argument lets one give a formula for rotation numbers associated to a hyperbolic structure on any noncompact hyperbolic surface in terms of Brooks functions on the associated (free) fundamental group. As before, let P be a fundamental domain for the surface, so that $D - P$ decomposes into regions on which the generators do ping-pong. Then each allowable pair xy of distinct letters in a reduced word moves up some fixed number n_{xy} of units. If $S_{g,p}$ is the surface of genus g with p punctures, then $\pi_1(S_{g,p})$ is free of rank $2g+p-1$ and we can take as generators $a_1, b_1, \cdots, a_g, b_g, c_1, \cdots, c_{p-1}$. We thereby obtain the following theorem.

THEOREM 4.76 (Rotation number formula). *Let $\mathcal{C}$ denote the following cyclically ordered set:*

$$\mathcal{C} = (a_1, b_1, a_1^{-1}, b_1^{-1}, \cdots, a_g, b_g, a_g^{-1}, b_g^{-1}, c_1, c_1^{-1}, \cdots, c_{p-1}, c_{p-1}^{-1})$$

For each pair x, y in $\mathcal{C}$ with $x \neq y, y^{-1}$ let m_{xy} be the integer $0 < m_{xy} < 4g+2p-2$ such that y is m_{xy} elements to the right of x in $\mathcal{C}$. Define

$$n_{xy} = \begin{cases} m_{xy} \text{ if } (y^{-1}, x, y) \text{ is positive in the circular order} \\ m_{xy} - (4g+2p-2) \text{ otherwise} \end{cases}$$

Then there is an equality

$$\text{rot}_{g,p} = \frac{1}{4g+2p-2}\left(\sum_{x \neq y \text{ or } y^{-1}} n_{xy}\overline{C}_{xy}\right)$$

where for each string σ, we let C_σ denote the counting function that counts copies of σ, and $\overline{C}_\sigma$ denotes its homogenization.

For example, let $S_{0,3}$ be the thrice punctured sphere, and let $\pi_1(S_{0,3}) = \langle a, b\rangle$ where a and b are loops around the punctures. Then if $\text{rot}_{0,3}$ denotes the homogeneous quasimorphism associated to the hyperbolic structure, there is a formula

$$\text{rot}_{0,3} = \frac{1}{2}\left(\overline{H}_{a^{-1}b} + \overline{H}_{ba^{-1}}\right)$$

Remark 4.77. In fact, the formula from Theorem 4.76 gives (after collecting terms)

$$\text{rot}_{0,3} = \frac{1}{4}\left(2\overline{H}_{a^{-1}b} + 2\overline{H}_{ba^{-1}} + \overline{C}_{ab} + \overline{C}_{b^{-1}a^{-1}} - \overline{C}_{a^{-1}b^{-1}} - \overline{C}_{ba}\right)$$

However, the function $C_{ab}+C_{b^{-1}a^{-1}}-C_{a^{-1}b^{-1}}-C_{ba}$ is uniformly bounded on any reduced word, as can be verified by a calculation, and therefore its homogenization is trivial.

Theorem 4.76 gives similar necessary and sufficient criteria in terms of counting quasimorphisms for geodesics in hyperbolic surfaces S corresponding to commutators in $\pi_1(S)$ to bound an immersed surface.

4.2.6. Rigidity Theorem. The content in the next few sections is taken largely from [**45**]. The main goal is to prove the following theorem:

Theorem 4.78 (Rigidity Theorem). *Let $F = \pi_1(S)$ where S is a compact oriented surface with $\chi(S) < 0$ and nonempty boundary.*

(1) *The projective class of the chain ∂S in $B_1^H(F)$ intersects the interior of a codimension one face π_S of the unit ball in the* scl *norm.*
(2) *The unique element of $Q(F)/H^1$ dual to π_S (up to scale) is the rotation quasimorphism associated to the action of $\pi_1(S)$ on the ideal boundary of the hyperbolic plane, coming from a hyperbolic structure on S.*

Theorem 4.78 reveals how surface topology and hyperbolic geometry are manifested in the bounded cohomology of a free group.

The proof is entirely elementary modulo Proposition 4.71, and depends only on constructing immersed surfaces in S with prescribed boundary. Technically, the result we prove is the following:

Theorem 4.79 (Immersion Theorem). *Let S be a compact oriented hyperbolic surface with (possibly empty) geodesic boundary. Let C be a homologically trivial rational chain, represented by a weighted union Γ of geodesics. Then for all sufficiently large N (depending on Γ), the chain $\Gamma + N\partial S$ virtually bounds a (positive) immersed surface.*

We show how to deduce Theorem 4.78 from Theorem 4.79.

PROOF. Let C be any rational chain in $B_1^H(F)$. By Theorem 4.79 and Proposition 4.71, for all sufficiently large N the chain $C + N\partial S$ in $B_1^H(F)$ satisfies

$$\text{scl}(C + N\partial S) = \text{rot}_S(C + N\partial S)/2$$

Hence the ray through ∂S intersects the interior of an edge of the unit ball of the scl norm restricted to the subspace $\langle C, \partial S\rangle$. Since C was arbitrary, the projective class of ∂S intersects the interior of a codimension one face π_S of the unit ball in the scl norm. By construction, this face is dual to rot_S (up to scale and H^1). □

REMARK 4.80. Proposition 4.71 says that a rational chain C virtually bounds a positive immersed surface in S if and only if $\text{scl}(C) = \text{rot}(C)/2$. By Theorem 4.78, this holds if and only if the projective class of C intersects the face π_S. If the support of C does not include ∂S, then $C - \epsilon\partial S$ cannot virtually bound a positive immersed surface in S for any positive ϵ. Consequently the projective class of such a C does not intersect the *interior* of π_S, but only its *boundary*.

REMARK 4.81. One still has a version of the Rigidity Theorem for closed surfaces. Let S be a closed, oriented hyperbolic surface. The hyperbolic structure lets us think of $\pi_1(S)$ as a subgroup of $\text{PSL}(2,\mathbb{R})$. Denote by G the preimage of this subgroup in $\widetilde{\text{SL}}(2,\mathbb{R})$. The group G is isomorphic to the fundamental group of the unit tangent bundle of S. There is a nontrivial central extension

$$\mathbb{Z} \to G \to \pi_1(S)$$

associated to the class of the generator of $H^2(S;\mathbb{Z})$. Let rot_Z denote the pullback of the rotation quasimorphism on $\widetilde{\text{SL}}(2,\mathbb{R})$ to G, and let Z denote the generator of the center of G. Theorem 4.79 and some elementary homological algebra implies that for any element $g \in [G,G]$, the quasimorphism rot_Z is extremal for $g + nZ$ whenever n is sufficiently large. Hence there is a codimension one face π_Z of the unit ball of the scl norm on $B_1^H(G)$, and the projective class of Z intersects the interior of this face.

By continuity, for any $g \in [G,G]$, the projective class of $g + nZ$ also intersects the interior of π_Z whenever n is sufficiently large (depending on g). Since Z is central, $\text{scl}(g + nZ + C) = \text{scl}(Z^n g + C)$ for any g and any chain C. Consequently, the projective class of the element $Z^n g$ also intersects the interior of π_Z whenever n is sufficiently large. Dually, rot_Z is the unique extremal homogeneous quasimorphism for $Z^n g$, up to scale and elements of H^1.

4.2.7. Proof of the immersion theorem. In this section we fix a surface S with $\pi_1(S) = F$ and a chain $C \in B_1^H(F)$ represented by a weighted sum of geodesics $\Gamma(C)$. Where there is no confusion, we abbreviate $\Gamma(C)$ to Γ. By LERF for surface groups (see Example 2.108) we can pass to a finite cover in which each component of the preimage of Γ is embedded (though of course the union will typically not be embedded). Let Γ' be the total (weighted) preimage of Γ in the cover S'. If Γ' cobounds a positively immersed surface with some multiple of $\partial S'$, this immersed surface projects to S and shows that the same is true of Γ. So without loss of generality, we can assume that every component of Γ is embedded.

If Γ_1 and Γ_2 virtually bound positive immersed surfaces, the same is true of $\Gamma_1 + \Gamma_2$, by Proposition 4.71 and the linearity of rot_S on B_1^H. The only homologically trivial chains in B_1^H represented by weighted sums of geodesics supported in ∂S are the multiples of ∂S, so to prove the theorem, it suffices to find any weighted collection of geodesics ∂ with support in ∂S so that $\Gamma + \partial$ virtually bounds a positive immersed surface. By abuse of notation, we say that $\Gamma + \partial$ virtually bounds a positive immersed surface if there is some (unspecified) ∂ with this property.

Suppose Γ_1 and Γ_2 are such that $\Gamma_1 - \Gamma_2 + \partial$ virtually bounds a positive immersed surface for some ∂. Let $i : T \to S$ be such an immersed surface. Then (again by LERF for surface groups) there are finite covers T', S' so that $i' : T' \to S'$ is an *embedding*. The difference $S' - i'(T')$ projects to S and shows that $\Gamma_2 - \Gamma_1 + \partial$ also virtually bounds a positive immersed surface (here ∂ typically stands for a different weighted collection of geodesics with support in ∂S). Define a relation $\sim$ on weighted collections of geodesics, where $\Gamma_1 \sim \Gamma_2$ if $\Gamma_1 - \Gamma_2 + \partial$ virtually bounds a positive immersed surface for some ∂ with support in ∂S. By the arguments above, this relation is reflexive, symmetric and transitive, and is consequently an *equivalence relation*. To prove the theorem therefore, we need only show that $\Gamma = \Gamma(C)$ satisfies $\Gamma \sim 0$.

LEMMA 4.82. *Let $S' \subset S$ be a subsurface with geodesic boundary, and let S'' be obtained from S' (topologically) by adding disks to close up some of the boundary components. Suppose that every boundary component of S' is either a boundary component of S, or is separating in S. Suppose further that γ and γ' are simple geodesics in S' that are homotopic in S''. Then $\gamma \sim \gamma'$.*

PROOF. Homotopic simple loops in S'' are isotopic in S''. Such an isotopy can be taken to be a sequence of simple moves which "push" γ over a single boundary component of S'. The result is realized at each stage by an embedded geodesic in S'. Every boundary component ∂_i of S' is either a boundary component of S, or is separating, and in either case $\partial_i \sim 0$. Hence $\gamma \sim \gamma'$ as claimed. □

Let δ be a family of pairwise disjoint essential separating geodesics which decompose S into a union of genus one subsurfaces S_i. There is a graph dual to this decomposition, with one vertex for each component of $S - \delta$, and one edge for each component of δ. Since each δ is separating, this dual graph is a tree. There are several possible such decompositions; for concreteness, choose a decomposition for which this dual graph is an interval. Note that a separating geodesic δ_i necessarily satisfies $\delta_i \sim 0$.

LEMMA 4.83. *Let γ be an embedded geodesic in S, and let δ as above separate S into genus 1 subsurfaces. Suppose γ intersects δ. Then there is an embedded geodesic 1-manifold γ' with at most two components, such that $\gamma \sim \gamma'$, and such that γ' intersects δ in fewer points than γ.*

PROOF. Every component of δ satisfies $\delta_i \sim 0$, so without loss of generality we can assume γ intersects δ transversely. There is at least one component S_i of $S - \delta$ such that γ intersects exactly one boundary component δ_i of S_i. Since δ_i is separating, the algebraic intersection number of γ with δ_i is zero, and therefore γ must intersect δ_i in at least two points with opposite signs. Let α be an arc of δ_i whose interior is disjoint from γ, and whose endpoints intersect γ with opposite signs. Build an embedded thrice punctured sphere in S by thickening γ, and attaching a 1-handle with core α. Isotope the boundary components of this thrice punctured sphere until they are (embedded, disjoint) geodesics. One component is γ; the other two components are γ'. □

By repeatedly applying Lemma 4.83, we can construct Γ' with $\Gamma \sim \Gamma'$, such that each geodesic in Γ' is embedded and contained in a genus one subsurface S' of S satisfying the hypothesis of Lemma 4.82. Let S'' be obtained from S' topologically by filling in all but one boundary component. Fix a standard basis

α, β of embedded geodesics in S' generating the homology of S''. Then γ represents $p\alpha + q\beta$ in homology. Since γ is embedded, p, q are coprime. We would like to show $\gamma \sim p\alpha + q\beta$. By induction, it suffices to show that the chain $a + b + a^{-1}b^{-1} \sim 0$ in a once-punctured torus, or equivalently that the chain $a + b + a^{-1}b^{-1} + [a, b]^n$ virtually bounds a positive immersed surface for some n. This can be proved by an explicit construction.

EXAMPLE 4.84. The chain $a + b + a^{-1}b^{-1} + [a, b]^2$ bounds an immersed surface in a once-punctured torus. One way to see this is to compute, using `scallop` to show $\text{scl}(a+b+a^{-1}b^{-1}+[a, b]^2) = 1$ and then verifying equality in Proposition 4.71, using the formula in Lemma 4.74 for rot. Another way is by explicit construction. There is an immersed four-holed sphere, found by Matthew Day, whose boundary is the chain $a + b + a^{-1}b^{-1} + [a, b]^2$. This surface is depicted in Figure 4.9 (compare with Figure 5 from [**45**]).

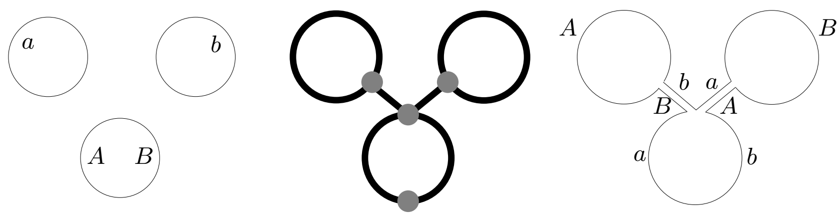


FIGURE 4.9. A 4-holed sphere that immerses in a once-punctured torus, with four boundary components (indicated by thin curves) in the conjugacy classes of a, b, $a^{-1}b^{-1}$ and $[a, b]^2$.

We now explain how to put these pieces together to prove the theorem.

PROOF. Let Γ in S be homologically trivial, with every component embedded. Decompose S along embedded separating geodesics δ as above into genus one subsurfaces. By Lemma 4.83, we can find Γ', a weighted sum of embedded geodesics, such that $\Gamma \sim \Gamma'$, and Γ' is disjoint from δ. For each component S' of $S - \delta$, let $\Gamma'(S')$ be the components of Γ' in S'. For each γ in $\Gamma'(S')$ there are coprime integers $p(\gamma)$ and $q(\gamma)$ so that $\gamma \sim p(\gamma)\alpha + q(\gamma)\beta$. But Γ is homologically trivial, and therefore the same is true of Γ' and $\Gamma'(S')$. Hence $\sum_\gamma p(\gamma) = \sum_\gamma q(\gamma) = 0$ and therefore $\Gamma'(S') \sim 0$. Since S' was arbitrary, $\Gamma' \sim 0$ and therefore $\Gamma \sim 0$. This completes the proof of Theorem 4.79 (and of Theorem 4.78). □

See [**45**] for more details and discussion.

4.2.8. Infinite dimensional faces. Theorem 4.78 can be "bootstrapped" in an interesting way. Let C be a rational chain in $B_1^H(F)$. The chain C is represented by a weighted collection Γ of geodesic loops in S where $\pi_1(S) = F$. By Theorem 4.24, there is an extremal surface T for C, i.e. a π_1-injective map $f : T, \partial T \to S, \Gamma$ realizing the infimum of $-\chi^-(T)/2n(T)$. Now, let C' be an arbitrary chain in $B_1^H(\pi_1(T))$, and Γ' a weighted collection of geodesic loops in T that it represents. By Theorem 4.78, for sufficiently large m the chain $\Gamma' + m\partial T$ virtually bounds an immersed surface. That is, there is an immersion $g : U, \partial U \to T, \Gamma' \cup \partial T$ for which $g(\partial U) = n'(\Gamma' + m\partial T)$ for some n'.

Lemma 4.85 (Bootstrap Lemma). *The surface* $f \circ g : U, \partial U \to S, f(\Gamma') \cup \Gamma$ *is extremal for some multiple of the chain* $f(C') + mC$ *in* $B_1^H(F)$.

Proof. By LERF, there are finite covers $T' \to T$ and $U' \to U$ so that g lifts to an embedding $g' : U' \to T'$. Clearly, it suffices to show that $f' \circ g'$ is extremal for some multiple of $f(C') + mC$. Since g' is an embedding, we can write T' as a union $T' = g'(U') \cup T''$. If $f' \circ g'$ is not extremal, there is some other $h : V, \partial V \to S, f(\Gamma') \cup \Gamma$ which is extremal for a (possibly different) multiple of $f(C') + mC$, and satisfies $-\chi^-(V)/2n(V) < -\chi(U')/2n(U')$. By the argument of Proposition 2.13, suitable covers of V and T'' can be glued up to produce a surface W which is extremal for Γ but satisfies $-\chi^-(W)/2n(W) < -\chi^-(T)/2n(T)$, contrary to the hypothesis that T is extremal. This contradiction shows that no such surface V exists, and therefore $f \circ g$ is extremal, as claimed. □

The following corollary is immediate:

Corollary 4.86. *Let* F *be a free group, and* $C \in B_1^H(F)$ *a rational chain. The projective class of* C *in* $B_1^H(F)$ *intersects the interior of an infinite dimensional face* π_C *of the unit ball in the* scl *norm. If* $f : \pi_1(T) \to F$ *is any extremal surface for* C*, then* $f_*(\pi_T) \to \pi_C$ *is isometric, in the sense that* $\text{scl}_{\pi_1(T)}(C') = \text{scl}_F(f_*(C'))$ *for all chains* C' *in the cone on* $\pi_T \subset B_1^H(\pi_1(T))$.

Proof. All that needs to be shown is that π_C is infinite dimensional, and to establish this it suffices to show that the image of $B_1^H(\pi_1(T))$ in $B_1^H(F)$ is infinite dimensional. Since T is extremal, $f_*(\pi_1(T))$ is a nontrivial finitely generated free subgroup of F. By Hall, free groups are virtual retracts (Example 2.107), so one can find infinitely many elements in $f_*(\pi_1(T))$ which are independent in $B_1^H(F)$. □

By convexity of the norm, the face π_C is well-defined. Note that Corollary 4.86 shows that extremal maps are norm-preserving on a nonempty open subset of B_1^H (compare with § 4.1.10).

Remark 4.87. Lemma 4.85 and Corollary 4.86 can also be deduced using quasimorphisms. Suppose $f : T \to S$ is extremal for some chain C. Let ϕ be an extremal quasimorphism for C with defect 1. Then $f^*\phi$ is an extremal quasimorphism for ∂T with defect 1 because T is extremal. By Theorem 4.78, $f^*\phi$ is equal to rot_T on $B_1^H(\pi_1(T))$. But rot_T is extremal on $B_1^H(\pi_1(T))$. Hence for every $C' \in \pi_T$, we have

$$\text{scl}_F(f_*(C')) \le \text{scl}_{\pi_1(T)}(C') = \text{rot}_T(C')/2 = \phi(f_*(C'))/2 \le \text{scl}_F(f_*(C'))$$

where the first inequality is monotonicity of scl, and the last inequality is Bavard duality.

One might wonder whether every face π_C has finite codimension. In fact, this is not the case. The following example is taken from [**45**].

Example 4.88. By Bavard duality, the codimension of π_C is one less than the dimension of the space of extremal quasimorphisms for C (mod H^1). Hence to exhibit a rational chain (in fact, an element of $[F, F]$) whose projective class intersects the interior of a face of infinite codimension, it suffices to exhibit a chain that admits an infinite dimensional space of extremal quasimorphisms.

Let $F = F_1 * F_2$ where F_1 and F_2 are both free of rank at least 2, and let $g \in [F_1, F_1]$ be nontrivial. Let $\phi_1 \in Q(F_1)$ be extremal for g, and let $\phi_2 \in Q(F_2)$ be arbitrary with $D(\phi_2) \le D(\phi_1)$. By the Hahn–Banach Theorem, there exists $\phi \in Q(F)$ that agrees with ϕ_i on F_i, and satisfies $D(\phi) = D(\phi_1)$.

EXAMPLE 4.89. Let ρ_t be a continuous family of (nonconjugate) indiscrete representations of F_2 into $\mathrm{PSL}(2,\mathbb{R})$. For a typical family ρ_t, the image $\rho_t(F_2)$ is dense in $\mathrm{PSL}(2,\mathbb{R})$ for all t, and therefore we can find (many) elements $g, h \in F_2$ generating a subgroup Γ so that $\rho_t(\Gamma)$ is discrete and purely hyperbolic, and the axes of $\rho_t(g)$ and $\rho_t(h)$ cross, for all t in some nontrivial interval I. Let rot_t be the homogeneous quasimorphism on F_2 (well-defined up to an element of H^1) associated to the representation ρ_t. Without loss of generality, we can choose rot_t to vary continuously as a function of t on every element of F_2. By construction, rot_t is an extremal quasimorphism for $[g,h]$, for all $t \in I$. On the other hand, for a suitable (indiscrete) family of representations ρ_t, for every nonempty interval I we can find a subinterval J, a point $p \in J$, and an element $f \in F_2$ for which $\mathrm{rot}_t(f)$ is elliptic for all $t \in J$ with $t < p$, and hyperbolic for all $t \in J$ with $t > p$. The quasimorphisms rot_t are constant on f for $t > p$ and nonconstant for $t < p$, so they span an infinite dimensional subspace of $Q(F_2)$. Hence the codimension of the face $\pi_{[g,h]}$ is infinite (compare with Burger–Iozzi [**30**]).

See [**45**] for more corollaries and discussion.

4.2.9. Discreteness of linear representations. Theorem 4.78 has applications to the study of symplectic representations of free and surface groups. For a basic reference to the theory of symplectic groups and representations, see [**31**] (we also return to this subject in more detail in § 5.2.3). We give a new proof of a relative version of rigidity theorems of [**93**] and [**31**], at least in an important special case. Roughly speaking, Goldman observed (in the case of $\mathrm{PSL}(2,\mathbb{R})$) that representations of surface groups of maximal Euler class are *discrete*. Burger–Iozzi–Wienhard extended this observation to symplectic groups, and characterized such representations geometrically.

The context is as follows. Let S be a compact oriented surface with boundary, and let $\rho : \pi_1(S) \to \mathrm{Sp}(2n,\mathbb{R})$ be a symplectic representation for which the conjugacy classes of boundary elements fix a Lagrangian subspace. This condition ensures that there is a well-defined relative Euler class (usually called the Maslov class for $n > 1$) which we denote $e_\rho \in H^2(S,\partial S;\mathbb{Z})$ associated to ρ (compare with § 4.2.4). The cohomology class e_ρ is bounded, with norm $n/2$, and therefore $|e_\rho([S])| \le -n\chi(S)$. A representation is said to be *maximal* (and e_ρ is *maximal*) if equality is achieved.

The following corollary says that maximal Zariski dense representations are discrete. We restrict to Zariski dense representations for simplicity; this condition is not necessary (see [**93, 31, 32**]).

COROLLARY 4.90 (Goldman, Burger–Iozzi–Wienhard). *Let S be a compact oriented surface with boundary. Let $\rho : \pi_1(S) \to \mathrm{Sp}(2n,\mathbb{R})$ be Zariski dense, and suppose that conjugacy classes of boundary elements fix a Lagrangian subspace. If e_ρ is maximal, ρ is discrete.*

PROOF. For the remainder of the proof, denote $\pi_1(S)$ by F and its commutator subgroup by F'. Since S has boundary, $e_\rho = [\delta\phi]$ where ϕ is in $Q(F)$, and is unique up to elements of H^1. For each $g \in F$, the value $\phi(g)$ (mod $\mathbb{Z}$) is the *symplectic rotation number*, and depends only on the image $\rho(g)$. Since e_ρ is maximal, ϕ is extremal for $\partial S \in B_1^H(F)$. Hence, by Theorem 4.78, it follows that the symplectic rotation number is *zero* on every $g \in F'$; in particular, $\rho(F')$ is not dense in

$\mathrm{Sp}(2n, \mathbb{R})$. Since $\mathrm{Sp}(2n, \mathbb{R})$ is simple, every Zariski dense subgroup is either discrete or dense (in the ordinary sense). If $\rho(F)$ is dense, then the closure of $\rho(F')$ is normal in $\mathrm{Sp}(2n, \mathbb{R})$. But $\mathrm{Sp}(2n, \mathbb{R})$ is simple, and the closure of $\rho(F')$ is a proper subgroup; hence $\rho(F)$ is discrete. □

Remark 4.91. The condition that boundary element fix Lagrangian subspaces is only included so that the Corollary can be phrased in terms of an integral Euler (Maslov) class. If $\rho : \pi_1(S) \to \mathrm{Sp}(2n, \mathbb{R})$ is *any* Zariski dense representation for which the pullback of the symplectic rotation quasimorphism (i.e. the quasimorphism ϕ above) is extremal for ∂S, then ∂S necessarily fixes a Lagrangian subspace.

4.2.10. Character Varieties. Any representation of a free group $\rho : F \to \mathrm{PSL}(2, \mathbb{R})$ lifts to $\widetilde{\mathrm{SL}}(2, \mathbb{R})$ and defines an associated homogeneous quasimorphism $\mathrm{rot} : F \to \mathbb{R}$ unique up to a homeomorphism. Given $a \in [F, F]$ one can ask what values this function can take as ρ varies over all homomorphisms.

We restrict attention to the case that F is free of rank 2, generated by elements a, b. The function rot only depends on the conjugacy class of ρ, and therefore we consider representations up to conjugacy. In fact, since ρ can typically be recovered just from the traces of elements, it makes sense to consider the *character variety*, consisting of the set of functions on F which are traces of some representation. For simplicity, it makes sense to study the $\mathrm{SL}(2, \mathbb{R})$ character variety instead, since traces are well defined there.

Definition 4.92. Let G be a finitely generated group. The *character variety* of G, denoted $X(G)$, is the set of functions $\chi : G \to \mathbb{R}$ for which $\chi = \mathrm{tr}(\rho)$ for some representation $\rho : G \to \mathrm{SL}(2, \mathbb{R})$.

Characters with representations in a fixed algebraic group satisfy many nontrivial (polynomial) relations, and a character is determined by its values on finitely many elements. This gives $X(G)$ the structure of a (real) algebraic variety. See [**60**] for an introduction to SL character varieties, and their applications to 3-manifolds.

Example 4.93. Let $G = F_2$, the free group on generators a, b. Since $\mathrm{SL}(2, \mathbb{R})$ is 3-dimensional, the space of $\mathrm{SL}(2, \mathbb{R})$ representations of F_2 is 6 dimensional, and the space of characters is 3-dimensional. If χ is a character, the co-ordinates $(x, y, z) = (\chi(a), \chi(b), \chi(ab))$ defines a map from $X(F_2)$ to $\mathbb{R}^3$. In fact, this map is an isomorphism onto the subset of $\mathbb{R}^3$ consisting of the union of the complement of the open cube $(-2, 2)^3$ together with the subset of triples inside the cube satisfying

$$x^2 + y^2 + z^2 - xyz \geq 4$$

Theorem 4.94. *Let $g \in [F_2, F_2]$. Then the set of values of $\mathrm{rot}(g)$ as one varies over all $\mathrm{SL}(2, \mathbb{R})$ representations of F_2 is a closed, connected interval, whose endpoints have the property that their image under $\cos(2\pi \cdot)$ is algebraic.*

Proof. Example 4.93 shows how to identify $X(F_2)$ with a semi-algebraic subset of $\mathbb{R}^3$. For every $g \in F_2$, the value of $\chi(g)$ is an integral polynomial in the values of $\chi(a), \chi(b), \chi(ab)$.

The function $\chi(g) : X(G) \to \mathbb{R}$ is therefore an integral polynomial on $\mathbb{R}^3$. An extremal value is a zero of a system of integral polynomial equations, and is therefore realized at an algebraic point. Since $2\cos(2\pi\ \mathrm{rot}(g)) = \chi(g)$, the result follows. □

Remark 4.95. A similar theorem can be proved with a similar proof with $\mathrm{Sp}(2n,\mathbb{R})$ or $\mathrm{SO}_0(n,2)$ in place of $\mathrm{SL}(2,\mathbb{R})$

4.3. Diagrams and small cancellation theory

The proof of Proposition 4.36 shows that in a fixed free group, every extremal surface can be built up from pieces (polygons and rectangles) of bounded complexity. A representation of a surface (with prescribed boundary) as a union of simple pieces drawn from some finite set is sometimes called a *diagram*. Diagrams can be represented graphically, and can be combined, composed and manipulated according to certain sets of rules. They have psychological value, as a way to represent algebraic information in geometric terms (e.g. as in Figure 4.9); and computational value. There are many different conventions for diagrams, depending on function and context.

Example 4.96. The conjugacy class $w = [a^2, b^2][a, b]$ has $\mathrm{scl} = 1$ in F_2. Let S be a hyperbolic once-punctured torus with basis a, b, and let γ be the geodesic associated to w. Then two copies of γ bound an immersed genus 2 surface T with two boundary components. Figure 4.10 depicts the surface T as a diagram,

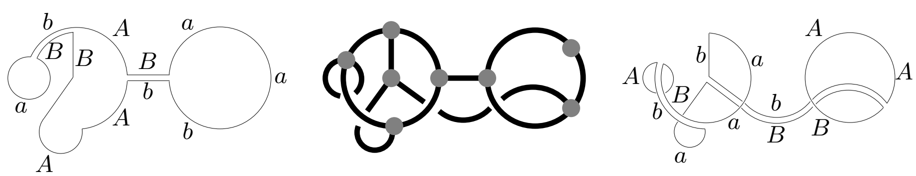


Figure 4.10. The surface T is obtained by thickening a graph with a cyclic ordering at the vertices. Edges of ∂T on opposite sides of each edge of the underlying graph are labeled by inverse elements of F_2. Each boundary component of ∂T is labeled by a cyclic conjugate of w.

obtained by thickening a graph whose vertices correspond to polygons, and edges to rectangles. The two copies of γ are indicated by thinner lines.

Remark 4.97. Any extremal surface obtained from the proof of Theorem 4.24 retracts in an obvious way to a graph with one edge for each rectangle, and one vertex for each polygon. To recover the surface (and therefore its boundary) from the graph, we need to specify a cyclic ordering of the edges at each vertex. A graph together with the choice of a cyclic ordering on the edges at each vertex is sometimes called a *ribbon graph* or a *fat graph*. Such objects appear in the study of dynamical systems, Hopf algebras, statistical mechanics, combinatorics, and many other fields; see [**17**].

4.3.1. Diagrams.

Diagrams (sometimes called *van Kampen diagrams*) were introduced by van Kampen in [**200**].

Let G be a group given by a presentation $G = \langle X \mid R\rangle$. Let F be the free group on X, and N the normal closure of R in X, so that $G = F/N$. The set R is said to have been *symmetrized* if all elements are cyclically reduced, and R is closed under taking cyclic permutations and inverses.

Definition 4.98. Let $w \in F$ be cyclically reduced. A *diagram* is a finite connected planar graph in which directed edges are labeled by elements of F, the boundary of each interior region is labeled by an element of R, and the boundary of the exterior region is labeled by a cyclic conjugate of w.

Since the graph associated to a diagram is assumed to be connected, interior regions are all homeomorphic to open disks. The boundary of a region is allowed to bump up against itself.

Note that the boundary label of a region depends on a choice of basepoint and a choice of orientation, or else the result differs by cyclic permutation or inverse. However, since R is symmetrized, membership in R is not affected by this ambiguity.

REMARK 4.99. A finite connected planar graph together with the regions it bounds is a simply-connected planar 2-complex. By abuse of notation we sometimes think of this 2-complex as the diagram.

If we assume elements of R are cyclically reduced, a map has no 1-valent vertices. Furthermore, if e_1, e_2 share a 2-valent vertex in common, we can replace $e_1 \cup e_2$ by $e_1 e_2$. Therefore in the sequel we assume every vertex is at least 3-valent.

LEMMA 4.100. *An element $w \in F$ admits a diagram if and only if it is in N.*

PROOF. There is a tautological cellular map from a diagram (thought of as a 2-complex) to a 2-complex associated to the presentation of G. Since the underlying 2-complex of a diagram is simply-connected, the boundary of the exterior region maps to a homotopically trivial loop. This exhibits w as an element of N.

Conversely, express w as a product of conjugates of elements of R. Denote this expression by a bunch of balloons in the plane tied by strings to a common basepoint, where each balloon is an element of R, and the string is the conjugating element. Then cancel adjacent edges whenever possible. The result is a finite connected planar graph whose boundary is a cyclically reduced word which is equal in F to w (after choosing a suitable basepoint and orientation), and therefore must be equal to w by uniqueness of reduced representatives in free groups. □

DEFINITION 4.101. A diagram is *reduced* if no two adjacent regions have boundaries which represent inverse elements of R, where the basepoint is taken to be some common vertex, and the orientations on the boundaries disagree (when compared with some orientation inherited from the plane).

Any diagram may be replaced by a reduced one, by collapsing nonreduced pairs of adjacent regions, thereby reducing the number of regions in the diagram until the process terminates.

DEFINITION 4.102. A word $b \in F$ is called a *piece* (relative to R) if there are distinct relations $ba_1, ba_2 \in R$.

An edge of a diagram between adjacent regions is a piece.

4.3.2. Small cancellation theory. In full generality, the theory of van Kampen diagrams is essentially combinatorial. However, when applied to groups with presentations that obey certain conditions (of a geometric nature), it makes contact with the theory of hyperbolic groups, negative curvature, regular languages, and so on. The geometric theory of diagrams arising from groups with presentations satisfying such conditions is called *small cancellation theory*.

Small cancellation theory has its origins in the work of Dehn [**63**], in which he posed the word and conjugacy problems for finitely presented groups, and solved these problems for fundamental groups of closed orientable 2-manifolds.

Dehn's insight was that surface groups have presentations with a single relator r with the property that for any cyclic conjugate s of r or r^{-1} with $s \neq r^{-1}$, the product sr has very little cancellation. Thus if a word in a surface group is trivial, it can be simplified immediately by finding a big subword consisting of more than half of some s.

It was not until the work of Lyndon [**138**] and Weinbaum [**203**] that the importance of geometry in Dehn's work was properly appreciated, and small cancellation theory began to be systematically applied to combinatorial group theory.

The hypotheses of small cancellation theory are conditions which a given symmetrized presentation might satisfy. Some of these conditions are as follows:

$C'(\lambda)$: every piece has length less than λ times the length of a relation it appears in.

$C(p)$: no relation is a product of fewer than p pieces. Equivalently, every region in a reduced diagram with no edges in common with the exterior region has at least p sides.

$T(q)$: any interior vertex in a reduced diagram has at least q incident edges.

Note that $C'(\lambda)$ implies $C(p)$ for $\lambda p < 1$.

Let D be a reduced diagram for an element $w \in G$. We can make D into a metric space by choosing a polygonal structure on each region and gluing these polygons together. The small cancellation conditions and the Gauss–Bonnet Theorem give upper bounds on the (distributional) curvature in D for a suitable choice of structure.

Example 4.103. Condition $C(6)$ implies that every polygon has at least 6 sides. Choose a metric for which each region is a constant curvature regular polygon with side lengths 1 and all angles $2\pi/3$. If a region has 6 sides, it will be a Euclidean hexagon with this metric. If it has more than 6 sides, it will be hyperbolic. At every 3-valent vertex these polygons fit together. At every vertex of valence more than 3, there is an "atom" of negative curvature. In particular, D with such a metric is locally *non-positively curved*, at least in the interior of D.

Similarly, condition $C(7)$ lets one construct a metric on D which is strictly negatively curved everywhere.

Remark 4.104. The local curvature conditions satisfied by D in Example 4.103 are sometimes expressed in terms of a (local) $\mathrm{CAT}(\kappa)$ *condition*, where $\kappa = 0$ under the hypothesis $C(6)$ (at least in the interior of D), and $\kappa = -1$ under the hypothesis $C(7)$. See [**24**] for a definition, and a discussion of the relationship between $\mathrm{CAT}(\kappa)$ and (δ-)hyperbolicity.

4.3.3. Diagrams on surfaces.

Schupp [**184**] generalized small cancellation theory to diagrams on closed surfaces.

Definition 4.105. Let Φ be a free group of countably infinite rank. A *quadratic word* in Φ is a word w in which every generator which occurs in w occurs exactly twice (possibly with opposite signs).

If we write this word on the boundary of a polygon, then after gluing edges in pairs we get a closed (orientable or non-orientable) surface. After composing with a suitable automorphism of Φ, the word w can be put in a canonical form

$$w = [a_1, b_1] \cdots [a_g, b_g]$$

if the resulting surface is orientable, or

$$w = a_1^2 \cdots a_g^2$$

otherwise, where each a_i, b_i is a generator in Φ.

Now let F be a free group on a generating set X, and let G be a quotient of F, given by some presentation $G = \langle X \mid R \rangle$. A *solution* of the equation $w = 1$ in G is a collection of words α_i, β_i in F for which the image of w under composition $\Phi \to F \to G$ sending each $a_i \to \alpha_i$ and $b_i \to \beta_i$, is trivial in G.

We restrict attention in what follows only to quadratic words that represent orientable surfaces. Let w be a quadratic word in Φ, and v a word in the generators of F representing 1 in G. Let D be a (planar) diagram whose boundary is v, corresponding to an expression of v as a product of conjugates of relations in R. After gluing up the boundary of D compatibly with w, we obtain a diagram on a closed orientable surface. This new diagram may not be reduced, because pairs of canceling regions which were not adjacent in D may now be adjacent in S. We can try to cancel regions which become adjacent in S as we did before; the result might cause the surface to undergo a compression in an essential simple closed curve, and we will obtain a finite set of simpler surfaces. It is possible that after finitely many such reductions, the entire surface is compressed away. This happens, for example, when the word v was already trivial in F. Schupp obtains a kind of converse:

THEOREM 4.106 (Schupp [**184**], Thm. 1). *Let w be an orientable quadratic word in Φ, and let v be a solution to $w = 1$ in $G = \langle X \mid R \rangle$. If v is nontrivial in $F = \langle X \rangle$, then there is a reduced diagram on an orientable surface defined by some endomorphic image of w.*

If the presentation of G satisfies suitable small cancellation conditions, one obtains an upper bound on the Euler characteristic of any surface containing a reduced diagram.

EXAMPLE 4.107 (Culler [**59**]). Let F be free on a set X, and let $g \in F$ be nontrivial and cyclically reduced. Let n be a positive integer and consider the group G_n with presentation $G_n = \langle X \mid g^n \rangle$.

Suppose some cyclic conjugate of g^{-1} shares a common initial word v of g of length more than 1/2 length(g). Write $g = vw$ and $g^{-1} = w^{-1}v^{-1}$. Since v is an initial word of some cyclic conjugate of g^{-1}, it is also a subword of g^{-2}. Since length(v) > length(w), there must be a nontrivial overlap of v and v^{-1}. Without loss of generality, $v = v_1v_2$ and $v^{-1} = v_2v_3$. By comparing lengths, $v_2 = v_2^{-1}$ which cannot happen in a free group.

Now, exhibit g^n as a product of commutators $g^n = [b_1, c_1] \cdots [b_m, c_m]$ in F. Let v (not the same v as above) be the (typically non-reduced) word in F obtained by concatenating words representing the b_i, c_i and their inverses. Notice that v is the image of an orientable quadratic word w in Φ. Schupp shows how to obtain a reduced surface diagram as follows. First start with a single planar region with boundary labeled by v. The word v is typically not cyclically reduced, so the boundary of the region can be inductively "folded" until the result is a *cactus*; i.e. a single innermost disk region with boundary labeled by g^n, and a forest attached to its outside boundary, so that the outer boundary is labeled by v. This cactus may be glued up according to the quadratic structure of w. The result is a "cactoid", i.e. a finite union of closed oriented surfaces and graphs. Throwing away the graph pieces, one obtains a surface of genus at most m, with a single tile whose boundary is labeled by g^n (for details, see [**184**], especially § 3).

Since the surface is oriented, the only pieces that appear correspond to common subwords in cyclic conjugates of g^n and g^{-n}. By the argument above, each such piece has length at most half of the length of g. Consequently we obtain a surface, and a tessellation on it containing one disk region with at least $2n$ edges, and with vertices each of valence at least 3. If we denote the number of faces, edges, vertices in the tessellation by f, e, v then $f = 1, e \geq n, v \leq 2e/3$. In other words, $\chi(S) \leq 1 - 2n/3$. On the other hand, the genus of S is at most m which can be taken to be equal to $\text{cl}(g^n)$. Taking $n \to \infty$, we obtain an estimate $\text{scl}(g) \geq 1/6$.

Remark 4.108. The methods of § 4.1, especially the proof of Theorem 4.24, gives another construction of a reduced surface. With notation as in the proof of Theorem 4.24, let $f : S, \partial S \to H, \gamma$ be a surface with one boundary component wrapping n times around γ, where γ is in the free homotopy class associated to a cyclically reduced word g. After compression and homotopy, the surface S is obtained by gluing rectangles and polygons. A decomposition of S as a union of rectangles and polygons determines a graph $\Gamma \subset S$ to which S deformation retracts, with one vertex for every polygon, and one edge for every rectangle (compare with Figure 4.10). One may obtain a reduced oriented surface diagram as a union $P \cup \Gamma$ where P is a disk whose boundary is labeled g^n.

Notice that one should *not* perform boundary compressions, but only compressions and homotopy. The reason is that boundary compressions might change the number of boundary components of S (though not the total degree with which they map to γ). So one can *not* apply the full power of the arguments of § 4.1 and assume that there is an *a priori* bound on the valence of the vertices (equivalent to a bound on the complexity of polygon types).

4.3.4. Right orderability. The lower bounds from the previous section can be improved by using *orderability* properties of free groups and their one-relator quotients. In fact a *sharp* lower bound on scl in free groups can be obtained along these lines, by the method of Duncan–Howie [**67**].

The proof depends on a well-known theorem of Brodskii:

Theorem 4.109 (Brodskii [**26**]). *Let F be a free group, and let g be a primitive element of $[F, F]$. Then the one-relator group $G := \langle F \mid g \rangle$ is right orderable.*

It also makes use of a Lemma of Howie:

Lemma 4.110 (Howie [**114**] Cor. 3.4). *Let $g \in F$ be primitive and cyclically reduced. Then no proper subword h of g represents the identity in $G := \langle F \mid g \rangle$.*

We are now in a position to obtain a sharp lower bound on scl in free groups. Duncan–Howie use the language of *reduced pictures*, which are very similar to Schupp's reduced diagrams (see § 4.3.3). The main theorem of Duncan–Howie, i.e. Theorem 3.3 [**67**], is an inequality about the combinatorics of such pictures, which implies the desired estimate on scl.

The argument given below is essentially a paraphrase of much of the material on pp. 229–233 of [**67**], with a few simplifications appropriate for our context.

Theorem 4.111 (Duncan–Howie [**67**], Thm 3.3). *Let F be a free group. Then* $\text{scl} \geq 1/2$ *for every nontrivial element.*

Proof. Free groups of every countable rank embed in the free group of rank 2, so by monotonicity of scl it suffices to prove the theorem in rank 2. Fix notation $F = \langle a, b \rangle$. Let g be an element of $[F, F]$. Since scl is characteristic, without loss of generality we take g to be cyclically reduced. Furthermore, we may assume that

g is not a proper power, since scl is multiplicative under powers. Since $g \in [F, F]$, the word length of g is at least 4, since both a and a^{-1} must appear in g with equal multiplicity, and similarly for b and b^{-1}.

Let $G = \langle F \mid g \rangle$. Fix an integer n, and let $G_n = \langle F \mid g^n \rangle$. There is a natural surjective homomorphism $G_n \to G$. Exhibit g^n as a product of commutators in F. As in Example 4.107 (also see Remark 4.108) we can find a reduced diagram on a surface S with $\text{genus}(S) \le \text{cl}(g^n)$, containing a single tile R. Let P be a polygonal disk mapping surjectively and cellularly onto R by $\varphi : P \to R$. We think of S as being obtained from P by gluing up edges in its boundary. The boundary of P is labeled by g^n.

Since g is cyclically reduced and primitive, there is a natural partition of ∂P into n copies of g. We label the vertices of ∂P by the image of the corresponding subword of g in G. In other words, if $|g| = m$, and if $\text{id} = g_0, g_1, \cdots, g_{m-1}$ are the proper prefixes of g, then each vertex of ∂P is labeled by an element $\bar{g}_i$ which is the image of g_i in G, where consecutive vertices are labeled $\bar{g}_i, \bar{g}_{i+1}$ with indices taken mod m. By Lemma 4.110, the $\bar{g}_i$ are all *distinct* for different values of i. Note that what is labeled is a vertex of P; each vertex in R is in the image of at least two vertices of P, and the labels are typically different.

Let σ be a piece in S, and let $\sigma^\pm$ be the two preimages in ∂P. The map φ gives an orientation-reversing identification of σ^+ and σ^-. If there is a vertex $v \in \sigma$ for which the preimages v^+, v^- in $\sigma^\pm$ have the same label $\bar{g}_i$, there is an adjacent vertex $w \in \sigma$ for which the preimages w^+, w^- get the labels $\bar{g}_{i+1}$ and $\bar{g}_{i-1}$ (labels taken mod m). But this means $\bar{g}_i^{-1}\bar{g}_{i+1} = \bar{g}_i^{-1}\bar{g}_{i-1}$ and therefore $\bar{g}_{i-1} = \bar{g}_{i+1}$. But $|g| \ge 4$ so this contradicts Lemma 4.110.

By Theorem 4.109, the group G is right orderable. Fix a right ordering $<$. If σ is a piece in S, we have seen that the labels of corresponding vertices in σ^+ and σ^- are all different. Let u and v be adjacent in σ, and $u^\pm, v^\pm$ the corresponding adjacent pairs of vertices in $\sigma^\pm$. Suppose u^+ has the label $\bar{g}_i$ and u^- has $\bar{g}_j$. Then (without loss of generality), v^+ has the label $\bar{g}_{i+1}$ and v^- has $\bar{g}_{j-1}$. Moreover,

$$x := \bar{g}_i^{-1}\bar{g}_{i+1} = \bar{g}_j^{-1}\bar{g}_{j-1}$$

by the defining property of (surface) diagrams. Since G is right orderable,

$$\bar{g}_i > \bar{g}_j \text{ if and only if } \bar{g}_{i+1} = \bar{g}_i x > \bar{g}_j x = \bar{g}_{j-1}$$

in other words, either the labels on vertices of σ^+ are all (unambiguously) *greater* than the labels on the corresponding vertices of σ^-, or they are all *less* than the labels on the corresponding vertices of σ^-. We may therefore unambiguously define a co-orientation on σ, pointing from the side corresponding to the edge in P with bigger labels, to the side corresponding to the edge in P with smaller labels.

Now, suppose v is a vertex at which at least three pieces meet. There are some finite collection v_i of preimages of v in ∂P. There is a connected graph Γ_v, whose vertices are the v_i, and whose edges correspond to pairs of points in the boundary of edges in ∂P that map to the same piece in S. Topologically, Γ_v is homeomorphic to a circle, which can be thought of as the link of the vertex v. The co-orientation on pieces determines an orientation on Γ_v. Since this orientation is compatible with the ordering on the labels of the v_i, there is no oriented cycle in Γ_v. If v_i is neither a source nor a sink, say that it is a *cusp*. Notice that for every vertex v, the graph Γ_v contains at least one source and one sink, so there are at least two v_i that are not cusps.

On the other hand, if g^+ and g^- are the highest and lowest labels which appear anywhere, then there are n vertices of ∂P labeled g^+ and n vertices labeled g^-, appearing in alternating order. The co-orientation on ∂P must change at least once between consecutive copies of g^+ and g^-, and therefore ∂P has at least $2n$ cusps.

Give P the structure of an ideal polygon, with an ideal vertex at each cusp. At every vertex v of the diagram, at least two of the preimages v_i are not cusps. If exactly two v_i are not cusps, then v is a smooth point. Otherwise, v has an atom of negative curvature of weight $(q-2)\pi$, where q is the number of v_i in the preimage of v which are not cusps. Since P has at least $2n$ ideal vertices, it has area at least $(2n-2)\pi$. Atoms of negative curvature reduce the area of S, so by Gauss–Bonnet, $\text{area}(P) = \text{area}(S) \le -2\pi\chi(S)$. Hence

$$(2n-2)\pi \le \text{area}(P) \le -2\pi\chi(S)$$

where $\chi(S) = 2 - 2 \cdot \text{genus}(S)$, and $\text{genus}(S) \le \text{cl}(g^n)$.

Rearranging this and taking the limit as $n \to \infty$ gives $\text{scl}(g) \ge 1/2$. □

Remark 4.112. Duncan–Howie state and prove their theorem in the more general context of an element g in a free product $A * B$ of locally indicable groups. The analogues of Theorem 4.109 and Lemma 4.110 are true for products of locally indicable groups, with essentially the same proofs.

Also compare with the discussion in § 2.7.5.

Corollary 4.113. *Let S be an orientable surface. Then* scl $\ge 1/2$ *for every nontrivial element of* $\pi_1(S)$.

Proof. If S is not closed, $\pi_1(S)$ is free, so this follows from Theorem 4.111. If S is closed of genus 0 or 1, every element is either trivial or essential in H_1, so scl is infinite for nontrivial elements. Closed surface groups of genus at least 2 are *residually free*; i.e. for any $a \in S$ there is a homomorphism to a free group $\varphi_a : \pi_1(S) \to F$ for which $\varphi_a(a)$ is nonzero (see e.g [**139**] for a proof). Since scl is monotone under homomorphisms, the corollary follows. □

4.3.5. An example. As explained in Remark 4.108, the construction of extremal surfaces from branched surfaces in § 4.1 can be reformulated in the language of surface diagrams. Let w be a cyclically reduced element of a free group F, and let S be a surface bounding some multiple of w, built from rectangles and polygons. Let T be the surface obtained from S by gluing in a disk to each boundary component. Then there is an associated diagram on T, whose edges are strings of consecutive rectangles and bigons in S, whose vertices are polygons in S with at least 3 ordinary arcs, and whose cells have boundaries which are labeled by finite powers of w.

We give an explicit construction of extremal surfaces for words of the form $[a,b][a,b^{-m}]$ for positive integers m. As asserted in Example 4.39, there is an equality

$$\text{scl}([a,b][a,b^{-m}]) = \frac{2m-3}{2m-2}$$

for $m \ge 2$. An inequality in one direction can be established by an explicit construction. In fact, for each $m \ge 2$ we will construct a genus $m-1$ surface with $2m-2$ boundary components, each of which wraps exactly once around $[a,b][a,b^{-m}]$. Hence there is a surface S with $-\chi^- = 4m-6$ and $n(S) = 2m-2$, so $\text{scl}([a,b][a,b^{-m}]) \le (2m-3)/(2m-2)$.

We begin by defining two tiles. The X tile has $b^{m-1}abA$ on the top, B^m on the bottom reading from left to right, and bA on the left, Ba on the right reading from top to bottom. The Y tile has b^m on the top, $B^{m-1}aBA$ on the bottom reading from left to right, and AB on the left, ab on the right reading from top to bottom. Note the X tile has $m+2$ letters on the top edge and m on the bottom edge, while the Y tile has m letters on the top edge, and $m+2$ on the bottom edge. See Figure 4.11.

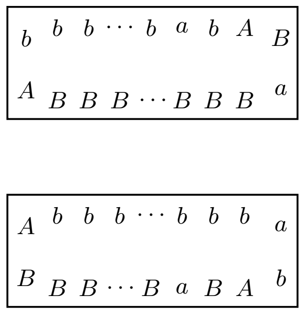


FIGURE 4.11. The tiles X and Y

Reading clockwise around each tile is a cyclic copy of the word $[a,b][a,b^{-m}]$. Tiles can be glued by gluing segments of their boundaries with opposite labels (where a and A are considered opposite labels, and similarly b and B). The left side of an X tile glues to the right side, and similarly the left side of a Y tile glues to the right side. Moreover, the bottom of an X tile glues to the top of a Y tile. Take $m-1$ copies of the X tile $XXXX\cdots X$ and glue left to right sides cyclically to make an annulus. Take a further $m-1$ copies of the Y tile $YYYY\cdots Y$ and glue left to right sides cyclically to make another annulus. Then glue the bottom of the X annulus to the top of the Y annulus to make a thicker annulus. The resulting labels, reading clockwise in each case, are $(b^{m-1}abA)^{m-1}$ on the top and $(B^{m-1}ABa)^{m-1}$ on the bottom. We glue these two components together in stages. At each stage, there are two boundary components, and we proceed to the next stage by gluing two disjoint segments in one component to disjoint segments in the other component with opposite labels. For clarity, let $n=m-1$ so that at the first stage the top is labeled $(b^n abA)^n$ and the bottom is labeled $(B^n ABa)^n$.

The result of gluing two segments in the top component to two segments in the bottom component has the effect of gluing on a four-times punctured sphere to the surface built so far. We indicate which segments are glued up at each step by using braces. The first two pairs of segments to be glued are $b^n \leftrightarrow B^n$ and $bAb \leftrightarrow BaB$:

$$(b^n abA)^{n-2}\,\underbrace{b^n}\,a\,\overbrace{bAb}\,b^{n-1}abA \text{ and } (B^n ABa)^{n-2}\,\underbrace{B^n}\,A\,\overbrace{BaB}\,B^{n-1}ABa$$

After gluing, this produces a new surface with two boundary components whose labels are

$$(b^n abA)^{n-2}Ab^{n-1}abA \text{ and } (B^n ABa)^{n-2}aB^{n-1}ABa$$

The next two pairs of segments to be glued are:

$$(b^n abA)^{n-3}\,\underbrace{b^n}\,a\,\overbrace{bAAb}\,b^{n-2}abA \text{ and } (B^n ABa)^{n-3}\,\underbrace{B^n}\,A\,\overbrace{BaaB}\,B^{n-2}ABa$$

which, after gluing, produces a new surface with two boundary components whose labels are

$$(b^n abA)^{n-3}Ab^{n-2}abA \text{ and } (B^n ABa)^{n-3}aB^{n-2}ABa$$

Proceed inductively, gluing up a b^n and a $bAAb$ in one boundary component to a B^n and a $BaaB$ in the other boundary component at each stage, until we are left with two boundary components labeled $AbabA$ and $aBABa$ which can be glued up completely. The final result is obtained from an annulus by attaching $n-1 = m-2$ pairs of 1-handles, and then gluing up a pair of circles at the end. The genus of the surface is therefore $m-1$. Moreover, it is tiled by $2m-2$ tiles, half of which are X tiles and half are Y tiles.

EXAMPLE 4.114. Let h denote the following linear combination of (small) counting quasimorphisms:

$$h = h_{abAB} + h_{aBBB} + h_{Abbb} + \frac{1}{2}(h_{bABa} + h_{ABaB} + h_{BaBB} + h_{BBBA} + h_{BBAb} + h_{BAbb} + h_{bbba} + h_{bbab} + h_{babA})$$

A (tedious) computation shows that $D(h) = 7/2$. It follows that $D(\overline{h}) \le 7$ for the homogenization $\overline{h}$. Moreover, $\overline{h}([a,b][a,b^{-m}]) = 15/2$ for all $m \ge 3$, so by Bavard duality we get a lower bound

$$\text{scl}([a,b][a,b^{-m}]) \ge 15/28 = 0.535714\ldots$$

We do not know whether a sharp lower bound can be achieved using counting quasimorphisms alone.

4.3.6. van Kampen soup, and thermodynamics of DNA. There is a curious diagrammatic relationship between scl and (a simplified model of) certain thermodynamic quantities associated to DNA (note that there is no suggestion that this model is physically realistic).

Deoxiribonucleic acid (DNA) is a nucleic acid that contains the genetic blueprint for all known living organisms. A molecule of DNA is a long polymer strand of simple units called *nucleotides*. The nucleotides in DNA (usually) come in four kinds, known as Adenine, Thymine, Guanine, and Cytosine (or A, T, G, C for short). Hence a molecule of DNA can be thought of as a (very) long string in this 4-letter alphabet, typically of length $\sim 10^8$.

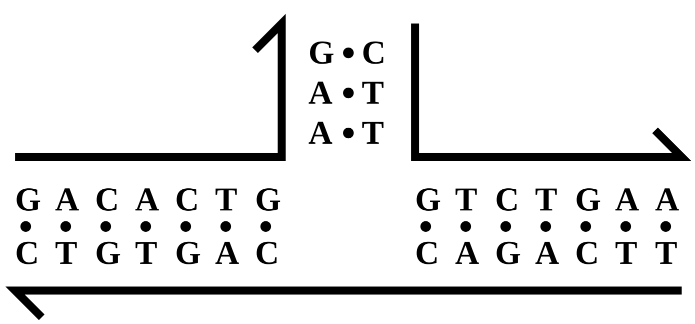


FIGURE 4.12. A 3-valent junction; figure adapted from [**186**]

These long strands tend to come in tightly bound oppositely aligned pairs, which match up nucleotides on the two molecules in complementary *base pairs.*

Each kind of nucleotide pairs with only one complementary kind: A with T, and C with G. The bonds joining base pairs are not covalent, and can be broken and rejoined easily.

Sometimes, "junctions" of three or more strands will form; see Figure 4.12. Three-valent junctions are the most common, but four-valent "Holliday junctions" can also form. There is an energy cost to forming such junctions, which in an idealization can be taken to be of order (valence $-\,2$), and is therefore proportional to $-\chi$. A reference for this material is [**186**].

Let $F = \langle a, b\rangle$ be the free group on two generators. A word in F can be "encoded" as a molecule of DNA by the encoding $a \to$ T, $a^{-1} \to$ A, $b \to$ C, and $b^{-1} \to$ G. If w is a cyclically reduced word in F, we can imagine preparing a "soup" of DNA containing many copies of the strand corresponding to $\dot{w} = \cdots www \cdots$. In thermodynamic equilibrium, the partition function has the form $Z = \sum_i e^{-E_i/k_B T}$ where k_B is Boltzmann's constant, T is temperature, and E_i is the energy of a configuration. At low temperature, minimal energy configurations tend to dominate; so $\mathrm{scl}(w)$ can be computed from the energy per unit volume of a van Kampen soup at low temperature.

CHAPTER 5

Irrationality and dynamics

The set of values of scl on all conjugacy classes in all finitely presented groups is a countable set. It is natural to try to characterize this set of real numbers, and to understand what kinds of arithmetic constraints exist on the values of scl in certain classes of groups.

As discussed in Chapter 4, the Rationality Theorem (i.e. Theorem 4.24) shows that for free groups (and more generally, for PQL groups) the scl norm is rational, and in particular, scl takes on values in $\mathbb{Q}$ in free groups. More generally, we saw that the unit ball of the scl norm on $B_1^H(F)$ is a rational polyhedron, and discussed the relationship of this example to the (polyhedral) Thurston norm on H_2 of an atoroidal irreducible 3-manifold.

It is natural to ask for which groups G the stable commutator length is rational on $[G, G]$. In fact, Gromov ([**99**], 6.C) explicitly asked whether scl is always rational, or at least algebraic, in general finitely presented groups. In the next section we describe an unexpected and elegant example due to Dongping Zhuang [**205**] of a finitely presented group in which the stable commutator length achieves *transcendental values*, thus answering Gromov's question in the negative.

There are two essential ingredients in Zhuang's examples: the groups he considers are *transformation groups* (i.e. groups of automorphisms of some geometric object), and they have an *arithmetic* origin. It is a general phenomenon, observed explicitly by Burger–Monod, Carter–Keller–Paige (as exposed by Dave Witte-Morris) and others, that (especially arithmetic) lattices in higher rank Lie groups generally admit no (nontrivial) quasimorphisms. On the other hand, such groups sometimes have nontrivial 2-dimensional bounded cohomology classes, which typically have a symplectic (or "causal") origin, which can be detected dynamically by realizing the groups as transformation groups. A central extension of such a group admits a nontrivial, but *finite dimensional* space of homogeneous quasimorphisms, and one may compute scl on such a group directly by Bavard duality, relating scl to dynamics.

In § 5.1 we discuss Zhuang's examples, which in some ways are the most elementary. In § 5.2 we discuss lattices in higher rank Lie groups from several different perspectives, eventually concentrating on lattices in symplectic groups as the most interesting examples. Finally, in § 5.3, we discuss some nonlinear generalizations of these ideas, which leads to the construction of quasimorphisms on braid groups and certain (low-dimensional) groups of area-preserving diffeomorphisms of surfaces. References for this chapter include [**28, 192, 205, 33, 34, 53, 159, 7, 86, 87**].

5.1. Stein–Thompson groups

In 1965, Richard Thompson [**195**] defined three groups $F \subset T \subset V$. Two of these (the groups T and V) were the first examples of finitely-presented, infinite simple groups. They can be defined as transformation groups (i.e. as groups of homeomorphisms of certain topological spaces): F is a group of homeomorphisms of an interval, T is a group of homeomorphisms of a circle, and V is a group of homeomorphisms of a Cantor set. Our interest in this section is on the groups F and T, and their generalizations. A basic reference for Thompson's groups is [**52**].

Definition 5.1. F is the group of orientation-preserving piecewise-linear (hereafter PL) homeomorphisms of the closed unit interval that are differentiable except at finitely many dyadic rational numbers (i.e. numbers of the form $p/2^q$ for integers p, q), and such that away from these discontinuities, the derivative is locally constant, and is equal to a power of 2.

T is the group of orientation-preserving PL homeomorphisms of the unit circle S^1 (thought of as $\mathbb{R}/\mathbb{Z}$) that maps dyadic rationals to dyadic rationals, has derivatives that are discontinuous at finitely many dyadic rationals, and are elsewhere equal to powers of 2.

Remark 5.2. All three groups can be defined as groups of *rotations* (in the sense of computer science) of infinite trivalent trees. In the case of F, the tree is rooted and planar; in the case of T, the tree is planar; in the case of V, the tree is neither rooted nor planar. See e.g. [**52**] § 2 or [**189**].

In this section we are interested in generalizations of the groups F and T due to Melanie Stein [**192**].

Definition 5.3. Let P be a multiplicative subgroup of the positive real numbers, and let A be a $\mathbb{Z}P$-submodule of the reals with $P \cdot A = A$. Choose a positive number $l \in A$. Define $F(l, A, P)$ to be the group of PL homeomorphisms of the interval $[0, l]$ taking $A \cap [0, l]$ to itself, whose derivatives have finitely many singularities in A, and take values in P.

Similarly, define $T(l, A, P)$ to be the group of PL homeomorphisms of the circle $\mathbb{R}/\langle l\rangle$ taking $A/\langle l\rangle$ to itself, whose derivatives have finitely many singularities in A, and take values in P.

Informally, we say that elements of $F(l, A, P)$ or $T(l, A, P)$ have *breakpoints* in A, and *slopes* in P.

Example 5.4. In this notation, Thompson's groups F and T are $F(1, \mathbb{Z}[\frac{1}{2}], \langle 2\rangle)$ and $T(1, \mathbb{Z}[\frac{1}{2}], \langle 2\rangle)$ respectively.

Stein showed in [**192**], following published and unpublished work of Brown [**28**], that for $l \in \mathbb{Z}$, for $A = \mathbb{Z}[1/n_1 n_2 \cdots n_k]$ and for $P = \langle n_1, \cdots, n_k\rangle$, the groups $F(l, A, P)$ and $T(l, A, P)$ are finitely presented, and in fact FP_∞ (i.e. there is a $K(G, 1)$ for these groups with only finitely many cells in each dimension). The method of proof is to explicitly find such a $K(G, 1)$. This is done by finding an action of these groups on suitable (explicitly described) contractible cubical complexes, such that the quotient complexes are homotopy equivalent to complexes with only finitely many cells in each dimension.

Example 5.5. A presentation for Thompson's group F is

$$F = \langle A, B \mid [AB^{-1}, A^{-1}BA], [AB^{-1}, A^{-2}BA^2]\rangle$$

A presentation for T is

$$T = \langle A, B, C \mid [AB^{-1}, A^{-1}BA], [AB^{-1}, A^{-2}BA^2], C^{-1}B(A^{-1}CB), \\ ((A^{-1}CB)(A^{-1}BA))^{-1}B(A^{-2}CB^2), (CA)^{-1}(A^{-1}CB)^2, C^3\rangle$$

These presentations are not terribly useful in practice, except that they do indicate algebraically how F is included as a subgroup of T. See [**52**], § 3 and § 5.

Zhuang's examples are central extensions of $T(l, A, P)$ for certain A and P as above. The remainder of this section is taken more or less verbatim from [**205**].

5.1.1. Factorization lemma. With notation as above, let $IP * A$ denote the submodule of A generated by elements of the form $(1-p)a$ where $a \in A$ and $p \in P$. In the sequel we sometimes abbreviate $T(l, A, P)$ by T for the sake of legibility (but T used in this sense should *not* be confused with Thompson's T).

LEMMA 5.6 (Stein [**192**]). *There is a natural homomorphism*

$$\nu : T(l, A, P) \to A/\langle IP * A, l\rangle$$

defined by $\nu(f) = f(a) - a$ for $f \in T$ and $a \in [0, l] \cap A$. If B denotes the kernel of ν, then $B' = T''$, the second commutator subgroup of T.

We use the following criterion of Bieri–Strebel (a proof appears in the appendix to [**192**]):

LEMMA 5.7 (Bieri–Strebel [**14**]). *Let $a, c, a', c' \in A$ with $a < c, a' < c'$. There is a* PL *homeomorphism of $\mathbb{R}$, with slopes in P and finitely many singularities in A, mapping $[a, c]$ onto $[a', c']$ iff $c' - a'$ is congruent to $c - a$ modulo $IP * A$.*

Lemma 5.6 and Lemma 5.7 together let one construct elements of T with desired properties. Let $f \in B$ be arbitrary. Zhuang proves the following factorization lemma.

LEMMA 5.8 (Zhuang [**205**], Lem. 3.4). *For any $f \in B$ there is a factorization $f = g_1 g_2$ in B where g_1 and g_2 both fix nonempty open arcs.*

PROOF. Note that any element which fixes a nonempty open arc fixes some point a in A, and is therefore in B by Lemma 5.6.

Let $f \in B$ be arbitrary. Choose points $a < b < a_1 < b_1 < c < d \in [0, l] \cap A$ such that $f([a, b]) = [a_1, b_1]$. Since $a_1 - a, b_1 - b \in IP * A$ (by the definition of B), Lemma 5.7 implies that there are PL homeomorphisms h_1, h_2 with slopes in P and singularities in A, sending $[b, c]$ to $[b_1, c]$ and $[d, a]$ to $[d, a_1]$ respectively. Now define

$$g = \begin{cases} f \text{ if } x \in [a, b] \\ h_1 \text{ if } x \in [b, c] \\ \text{id if } x \in [c, d] \\ h_2 \text{ if } x \in [d, a] \end{cases}$$

Set $g_1 = fg^{-1}$ and $g_2 = g$. Then $f = g_1 g_2$, and both g_1 and g_2 fix nonempty open arcs. □

REMARK 5.9. Factorization or "fragmentation" lemmas, together with Mayer–Vietoris and Künneth formulae, are generally the key to computing the (bounded co-) homology of transformation groups. Such techniques are used pervasively in the theory of foliations; see e.g. Tsuboi's survey [**199**].

For each $\theta \in IP * A$ the rotation R_θ is in B. The set of such θ is dense in $[0, l]$. So for $i = 1, 2$, let g_i be as in Lemma 5.8, and choose θ_i so that $R_{\theta_i} \in B$, and $h_i := R_{\theta_i} g_i R_{\theta_i}^{-1}$ has support contained in $(0, l)$.

5.1.2. Calculation of commutator subgroup. Let $F(l,A,P)$ denote the subgroup of $T(l,A,P)$ fixing 0. We abbreviate $F(l,A,P)$ by F, and think of F as a group of PL homeomorphisms of the interval $[0,l]$. Notice that $F \subset B$. There is a natural homomorphism

$$\rho : F \to P \times P$$

defined by $\rho(f) = (f'(0+), f'(l-))$; i.e. the image of ρ is the pair of elements of P consisting of the derivative of f at 0 from the right, and the derivative of f at l from the left. Let $B_1 = \ker \rho$. Note that $h_1, h_2 \in B_1$, since their support is contained strictly in the interior of $[0,l]$.

Theorem 5.10 (Stein [**192**]). *With notation as above, the commutator subgroup B_1' is simple, and $B_1' = F'$.*

On the other hand, one has the following theorem of Brown (see [**192**] for a proof):

Theorem 5.11 (Brown). *With notation as above, there is an isomorphism*

$$H_*(F) \cong H_*(B_1) \otimes H_*(P \times P)$$

We now specialize to the case that $l = 1, A = \mathbb{Z}[\frac{1}{pq}], P = \langle p, q \rangle$. Here p and q are arbitrary integers which form a basis for $\langle p, q \rangle$ (this is satisfied for example if p and q are distinct primes). We write $T_{p,q}, F_{p,q}$ for $T(l,A,P), F(l,A,P)$ in this case.

In [**192**], Stein explicitly calculates the homology of such $F_{p,q}$.

Lemma 5.12 (Stein, [**192**] Thm. 4.7). *With notation as above, $H_1(F_{p,q})$ is free Abelian with rank $2(d+1)$ where d is the greatest common divisor of $p-1$ and $q-1$.*

If $d = 1$ (for instance if $p = 2, q = 3$), Lemma 5.12 implies that $H_1(F_{p,q}) = \mathbb{Z}^4 = H_1(P \times P)$. Theorem 5.11 therefore implies that $H_1(B_1) = 1$ and therefore $B_1 = B_1' = F_{p,q}'$. By Lemma 5.8 and the definition of the h_i, we see that every element of B can be written as a product of conjugates of commutators in $B_1 \subset B$. In particular, B is perfect.

By Lemma 5.6, $B = B' = T_{p,q}''$. Since $T_{p,q}' \subset B$ (because B is the kernel of ν, which is a map from T to an Abelian group) we get $B = T_{p,q}'$. Furthermore, when $l = 1$ and $d = 1$, the submodule $\langle IP * A, 1 \rangle$ is actually equal to A, so ν is the zero map. Hence $T_{p,q}$ is perfect in this case.

5.1.3. Calculation of scl. The final ingredient we need is the following:

Theorem 5.13 (Calegari [**41**], Thm. A). *Let G be a subgroup of $\mathrm{PL}^+(I)$. Then* scl *vanishes on $[G,G]$.*

Proof. Let $g \in [G,G]$, and let H be a finitely generated subgroup so that $g \in [H,H]$. The fixed point set of any element of $\mathrm{PL}^+(I)$ is a finite union of points and closed intervals, so the same is true for the common fixed point set of a finitely generated group. Let $\mathrm{fix}(H)$ denote this common fixed point set, and enumerate the (finitely many) complementary open intervals as $I_1, I_2, \cdots, I_m$.

For each interval I_j there is a homomorphism $\rho_j : H \to \mathbb{R} \oplus \mathbb{R}$ defined by $\rho_j(h) = (\log dh^+(I_j^-), \log dh^-(I_j^+))$ where I_j^+ denotes the positive endpoint of the interval I_j, and I_j^- denotes the negative endpoint, and dh^+, dh^- denotes derivative from the right and from the left respectively. Let $\rho : H \to \mathbb{R}^{2m}$ be the direct sum of these homomorphisms, and let H_0 denote the kernel. Suppose $h \in [H_0, H_0]$, and let

K be a finitely generated subgroup of H_0 with $h \in [K,K]$. Then $\text{fix}(K)$ contains a neighborhood of each endpoint of each interval I_j, so there are closed intervals I'_j contained in the interior of the I_j such that the support of K is contained in the union of the I'_j.

For each closed interval J contained in the interior of some I_i, there is $j \in H$ with $j(J) \cap J = \emptyset$. By replacing j by its inverse if necessary, there is such a j which moves J to the right. We assume by induction that for any set of intervals J_i closed in the interior of each I_i, there is $j \in H$ with $j(J_i) \cap J_i = \emptyset$ for all $1 \le i \le r$. Let k satisfy $k(J_{r+1}) \cap J_{r+1} = \emptyset$ and k moves J_{r+1} to the right. Let J'_i be the smallest closed interval in the interior of I_i containing $J_i \cup k(J_i)$. By the induction hypothesis there is $j' \in H$ with $j'(J'_i) \cap J'_i = \emptyset$ for $1 \le i \le r$. Replacing j' by its inverse if necessary, we may further assume that j' moves the leftmost point of J_{r+1} to the right. Then $j'k(J_i) \cap J_i = 0$ for $1 \le i \le r+1$. It follows that we can find a single element $j \in H$ such that $j(I'_i) \cap I'_i = \emptyset$ for all i simultaneously.

For any n there is an injection $\Delta_n : K \to H$ defined by

$$\Delta_n(c) = \prod_{i=0}^{n} c^{j^i}$$

where j is as above, and the superscript denotes conjugation. Define

$$h' = \prod_{i=0}^{n} (h^{i+1})^{j^i}$$

Then $[h', j] = \Delta_n(h)(h^{-n-1})^{j^{n+1}}$. On the other hand, if $h = [a_1, b_1][a_2, b_2] \cdots [a_s, b_s]$ with $a_i, b_i \in K$ then $\Delta_n(h) = [\Delta_n(a_1), \Delta_n(b_1)] \cdots [\Delta_n(a_s), \Delta_n(b_s)]$. It follows that $\text{cl}(h^{n+1}) \le s+1$ in H and therefore $\text{scl}(h) = 0$, also in H. Since $h \in [H_0, H_0]$ was arbitrary, it follows that scl in H vanishes identically on $[H_0, H_0]$. On the other hand, $H/[H_0, H_0]$ is two-step solvable, and therefore amenable. Since scl vanishes in the commutator subgroup of an amenable group, for every element $g \in [H, H]$ there is a power n such that

$$g^n = [a_1, b_1] \cdots [a_s, b_s]c$$

where s/n is as small as we like, and $c \in [H_0, H_0]$. If ϕ is a homogeneous quasimorphism on H of defect 1, then ϕ vanishes on c, and therefore has value $\le 2s$ on g^n. Hence $\text{scl}(g) = 0$ in H, and therefore also in G. Since $g \in [G, G]$ was arbitrary, the theorem is proved. □

Remark 5.14. Notice the use of the Münchhausen trick (i.e. Example 3.66) in the construction of Δ_n.

We are now in a position to determine scl in $T_{p,q}$.

Lemma 5.15 (Zhuang [**205**], Lem. 3.8). *Let $T_{p,q}$ be as above where $d = \gcd(p-1, q-1) = 1$. Then* scl *vanishes on $T'_{p,q} = T_{p,q}$.*

Proof. Let ϕ be a homogeneous quasimorphism on $T_{p,q}$, and let $f \in T_{p,q}$ be arbitrary. By Lemma 5.8 we can write $f = g_1 g_2$ and $h_i = R_{\theta_i} g_i R_{\theta_i}^{-1}$ where each $h_i \in B_1$. Since B_1 is a perfect subgroup of $\text{PL}^+(I)$, Theorem 5.13 implies that $\text{scl}(h_i) = 0$ in B_1. Note that ϕ restricts to a homogeneous quasimorphism on B_1, and therefore by Bavard's Duality Theorem 2.70 we have $\phi(h_i) = 0$, and therefore $\phi(g_i) = 0$. But $f = g_1 g_2$, so

$$|\phi(f)| \le D(\phi)$$

by the definition of the defect. Since f was arbitrary, ϕ is uniformly bounded on $T_{p,q}$. A bounded homogeneous quasimorphism is identically zero. Since ϕ was arbitrary, scl is identically zero on $T_{p,q}$ by another application of Bavard's Duality Theorem. □

There is a natural central extension

$$0 \to \mathbb{Z} \to \widehat{T}_{p,q} \to T_{p,q} \to 0$$

where $\widehat{T}_{p,q}$ is the subgroup of $\text{Homeo}^+(\mathbb{R})$ which cover elements of $T_{p,q}$ under the covering projection $\mathbb{R} \to S^1$. Note that $\widehat{T}_{p,q}$ is finitely presented, since $T_{p,q}$ is. The class of this central extension is the *Euler class* of the natural action of $T_{p,q}$ on S^1. Since $\mathbb{Z}$ is amenable, Theorem 2.49 shows that the exact sequence induces an isomorphism $H_b^2(T_{p,q};\mathbb{R}) \to H_b^2(\widehat{T}_{p,q};\mathbb{R})$.

On the other hand, by construction, the kernel of the map in ordinary cohomology $H^2(T_{p,q};\mathbb{R}) \to H^2(\widehat{T}_{p,q};\mathbb{R})$ is 1-dimensional, generated by the Euler class. The usual five term exact sequence in cohomology for an extension (i.e. the Hochschild–Serre sequence; see § 1.1.6) implies that $H^1(\widehat{T}_{p,q};\mathbb{R})$ vanishes. By Theorem 2.50 the space $Q(\widehat{T}_{p,q})$ is 1-dimensional, and generated by rotation number, as in § 2.3.3. As in Proposition 2.92, $D(\text{rot}) = 1$. By Bavard's Duality Theorem we have the following:

THEOREM 5.16 (Zhuang [**205**], Thm. 3.9). *With notation as above, and for p, q satisfying $\gcd(p-1, q-1) = 1$, for any element $f \in \widehat{T}_{p,q}$ there is an equality*

$$\text{scl}(f) = \frac{|\text{rot}(f)|}{2}$$

We will see more examples of such an intimate relationship between scl and dynamics in the sequel.

5.1.4. Rotation numbers in Stein–Thompson groups. Rotation numbers in Stein–Thompson groups have been well-studied by Isabelle Liousse [**137**]. She proves the following:

THEOREM 5.17 (Liousse [**137**], Thm. 2.C′). *Any number of the form $\frac{\log\alpha}{\log\beta}$ mod $\mathbb{Z}$ where $\alpha, \beta \in \langle p, q\rangle$ can be realized as the rotation number of an element of the group $T(d, \mathbb{Z}[\frac{1}{pq}], \langle p, q\rangle)$ where $d = \gcd(p-1, q-1)$.*

For concreteness, take $p = 2, q = 3$. An example is the following:

EXAMPLE 5.18 (Liousse [**137**]). Define $a \in T_{2,3}$ by

$$a = \begin{cases} \frac{2}{3}x + \frac{2}{3} \text{ if } x \in [0, \frac{1}{2}] \\ \frac{4}{3}x - \frac{2}{3} \text{ if } x \in [\frac{1}{2}, 1] \end{cases}$$

Then any lift $\widehat{a}$ of a to $\widehat{T}_{2,3}$ has rotation number $\frac{\log 3}{\log 2}$ (mod $\mathbb{Z}$), and consequently $\text{scl}(\widehat{a})$ is irrational in $\widehat{T}_{2,3}$. In fact, scl in this case is transcendental, by the celebrated theorem of Gelfond and Schneider ([**89**],[**183**]). The graph of a is illustrated in Figure 5.1.

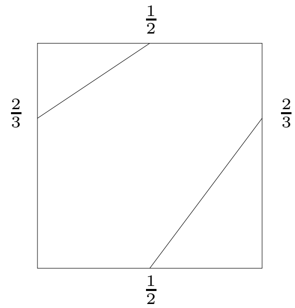


FIGURE 5.1. Graph of the homeomorphism $a \in T_{2,3}$

The map $h : x \to 2 - 2^{1-x}$ for $x \in [0, 1]$ conjugates a to a rigid rotation by $\log 3/\log 2$. This example is very closely related to examples studied also by Boshernitzan [**18**]. For a full discussion, and an explanation of this and related phenomena, see Liousse [**137**], § 3.

COROLLARY 5.19 (Zhuang [**205**]). *There exists a finitely presented group containing elements with transcendental* scl.

This answers in the negative question (c) in Gromov [**99**], page 142.

REMARK 5.20. Work of Ghys–Sergiescu [**92**] already shows that the classical Thompson group T is uniformly perfect, and therefore its central extension $\widehat{T}$ satisfies $\dim(Q(\widehat{T})) = 1$, spanned by rotation number. However, [**92**] show that every element of T has a periodic point in S^1, and therefore rotation number (and consequently scl) is rational in $\widehat{T}$. In any case, $\widehat{T}$ is an example of a finitely presented group whose scl spectrum is exactly equal to the non-negative rational numbers.

5.2. Groups with few quasimorphisms

The examples in § 5.1 suggest that it is fruitful to study examples of groups with H_b^2 finite dimensional. If G is a finitely presented group with scl identically zero, then $H_b^2(G)$ injects into the finite dimensional space $H^2(G)$ by Theorem 2.50. If $\widehat{G}$ is a central extension of G, then $Q(\widehat{G})$ is finite dimensional, and scl in $\widehat{G}$ can be computed by Bavard duality. The Stein–Thompson groups discussed in § 5.1 are examples of this kind. It is psychologically useful to think of such groups as "lattices" (in a certain sense) in the group of PL homeomorphisms of S^1. Thinking of these groups in this way connects them to a wider class of examples which we now discuss.

5.2.1. Higher rank lattices. The main references for this section are [**33, 34**] and [**66**]. Using tools from the theory of continuous bounded cohomology (see [**157**]), Burger–Monod show that the natural map from bounded cohomology to ordinary cohomology in dimension 2 is injective for a large class of important groups, namely lattices in higher rank Lie groups.

The main theorems of [**33, 34**] are stated in very general terms; we state these theorems for lattices in real Lie groups, for simplicity. First we recall some definitions.

DEFINITION 5.21. Let G be a closed subgroup of $\mathrm{SL}(m, \mathbb{R})$ for some m. A closed, connected subgroup T of G is a *torus* if T is diagonalizable over $\mathbb{C}$; i.e. if there is $g \in \mathrm{GL}(m, \mathbb{C})$ such that $g^{-1}Tg$ consists entirely of diagonal matrices. A torus T in G is $\mathbb{R}$-*split* if T is diagonalizable over $\mathbb{R}$; i.e. if there is $g \in \mathrm{GL}(m, \mathbb{R})$ such that $g^{-1}Tg$ consists entirely of diagonal matrices.

EXAMPLE 5.22. The subgroup $\mathrm{SO}(2, \mathbb{R})$ in $\mathrm{SL}(2, \mathbb{R})$ is a torus, but not an $\mathbb{R}$-split torus, since the eigenvalues of most elements are not real. On the other hand, the subgroup consisting of matrices of the form $\left(\begin{smallmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{smallmatrix}\right)$ where $\lambda \in \mathbb{R}^*$ is a (maximal) $\mathbb{R}$-split torus.

For G a real Lie group (not necessarily a matrix group), a closed, connected subgroup T is an $\mathbb{R}$-split torus if for every $x \in T$, the conjugation action of x on the Lie algebra of G is diagonalizable, with all real eigenvalues.

DEFINITION 5.23. Let G be a real Lie group. The *real rank* of G, denoted $\mathrm{rank}_{\mathbb{R}} G$, is the dimension of any maximal $\mathbb{R}$-split torus of G.

DEFINITION 5.24. A Lie group is said to be *simple* if it has no nontrivial, closed, proper, normal subgroups, and is not Abelian. It is *almost simple* if the only closed, proper, normal subgroups are finite.

REMARK 5.25. With this definition, the Lie group $\mathrm{SL}(2, \mathbb{R})$ is almost simple, since the only closed proper normal subgroup is the center $\pm\mathrm{id}$, but its universal cover $\widetilde{\mathrm{SL}}(2, \mathbb{R})$ is not almost simple, since its center is $\mathbb{Z}$.

A *lattice* Γ in a Lie group G is a discrete subgroup such that $\Gamma \backslash G$ has finite volume. A lattice is *uniform* (or *cocompact*) if $\Gamma \backslash G$ is compact, and *nonuniform* otherwise. A lattice Γ in a Lie group which is a nontrivial product $G = \prod_a G_a$ is *irreducible* if the projection of Γ to each proper product of factors is dense.

The following

THEOREM 5.26 (Burger–Monod [**34**], Thm. 21, Cor. 24). *Let Γ be an irreducible lattice in a finite product $G = \prod_a G_a$ where G_a are connected, almost-simple non-compact real Lie groups. If*

$$\sum_{a \in A} \mathrm{rank}_{\mathbb{R}} G_a \geq 2$$

then $H_b^2(\Gamma; \mathbb{R}) \to H^2(\Gamma; \mathbb{R})$ is injective.

REMARK 5.27. When Γ as above is uniform, this is contained in Theorem 1.1 from [**33**].

EXAMPLE 5.28. As an example we can take $G = \mathrm{SL}(2, \mathbb{R}) \times \mathrm{SL}(2, \mathbb{R})$. There is a well-known construction of lattices in $\mathrm{SL}(2, \mathbb{R}) \times \mathrm{SL}(2, \mathbb{R})$ using quaternion algebras, which we now describe. A standard reference for this material is Vignéras [**202**].

Let F be a number field (i.e. a finite algebraic extension of $\mathbb{Q}$), all of whose embeddings in $\mathbb{C}$ are contained in the real numbers. Such a field is said to be *totally real* and can be obtained, for instance, by taking a polynomial with rational coefficients all of whose roots are real, and adjoining to $\mathbb{Q}$ all of these roots. A *quaternion algebra* A over F is an algebra which as a group is a 4-dimensional vector space over F generated by elements $1, i, j, k$ with an associative and distributive multiplication law satisfying $i^2 = a, j^2 = b, k = ij = -ji$ for some $a, b \in F$. Such an algebra is typically denoted

$$A = \left(\frac{a, b}{F} \right)$$

A (Galois) embedding of F into $\mathbb{R}$ induces an inclusion of A into a quaternion algebra over $\mathbb{R}$. The only two such algebras, up to isomorphism are the matrix algebra $M_2(\mathbb{R})$, and the ring of Hamilton's quaternions $\mathbb{H}$. An embedding $\sigma : F \to \mathbb{R}$ is *ramified* if $A \otimes_{\sigma F} \mathbb{R} \cong \mathbb{H}$. Let $\mathcal{O}_F$ denote the ring of algebraic integers in F. It is finitely generated over $\mathbb{Z}$. An *order* $\mathcal{O}$ in A is a subring of A containing 1 that generates A over F, and is a finitely generated $\mathcal{O}_F$-module. If $x = x_0+x_1i+x_2j+x_3k$ is an arbitrary element of A, where the $x_i \in F$, the *norm* of x is $x_0^2-x_1^2a-x_2^2b+x_3^2ab$ and the *trace* is $2x_0$. The norm is a multiplicative homormorphism from A to F. If $\mathcal{O}$ is an order in A, the elements $\mathcal{O}^1$ of norm 1 are a group under multiplication.

Suppose that A is ramified at all but exactly two real embeddings of F. Consider the diagonal embedding

$$\rho : A \to M_2(\mathbb{R}) \times M_2(\mathbb{R}) \times \mathbb{H} \times \cdots \times \mathbb{H}$$

where each term is the embedding of A into $A \otimes_{\sigma_i F} \mathbb{R}$ associated to an embedding $\sigma_i : F \to \mathbb{R}$.

THEOREM 5.29. *With notation as above, the image* $\Gamma := \rho(\mathcal{O}^1)$ *is an irreducible lattice in the product*

$$\rho(\mathcal{O}^1) \subset \mathrm{SL}(2,\mathbb{R}) \times \mathrm{SL}(2,\mathbb{R}) \times \mathrm{SU}(2) \times \cdots \times \mathrm{SU}(2)$$

Moreover, if the degree of F *is at least* 3*, the lattice* Γ *is uniform.*

See e.g. [**202**] for a proof. Since the SU(2) factors are all compact, the image of Γ in $\mathrm{SL}(2,\mathbb{R}) \times \mathrm{SL}(2,\mathbb{R})$ is also a lattice.

If $x \in A$ and $\sigma : F \to \mathbb{R}$ is an unramified embedding inducing $\rho_\sigma : A \to M_2(\mathbb{R})$, the trace of the matrix $\rho_\sigma(x)$ is equal to the image under σ of the trace of x. In particular, these traces are *algebraic* numbers, contained in $\sigma(F)$. If $x \in \mathcal{O}^1$ and $g = \sigma(x) \in \mathrm{SL}(2,\mathbb{R})$, we can think of $\mathrm{SL}(2,\mathbb{R})$ acting on a circle, factoring through $\mathrm{SL}(2,\mathbb{R}) \to \mathrm{PSL}(2,\mathbb{R})$. The rotation number of g under this action is

$$\mathrm{rot}(g) = \frac{\cos^{-1}(\mathrm{trace}(g)/2)}{\pi}$$

mod $\mathbb{Z}$, providing $|\mathrm{trace}(g)| \le 2$. By Gelfond–Schneider, these rotation numbers are transcendental when they are not rational. Moreover, they are rational for only finitely many conjugacy classes in $\mathcal{O}^1$.

Let Γ be such a lattice, and consider the preimage $\widehat{\Gamma}$ in $\mathrm{SL}(2,\mathbb{R}) \times \widetilde{\mathrm{SL}}(2,\mathbb{R})$. The group Γ is finitely presented, since it has a compact fundamental domain for its action on the contractible space $\mathbb{H}^2 \times \mathbb{H}^2$. Since $\widehat{\Gamma}$ is a central $\mathbb{Z}$ extension, it is also finitely presented. As in § 5.1 the group $Q(\widehat{\Gamma})$ is one dimensional, generated by rotation number on the $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ factor. Hence for $g \in \widehat{\Gamma}$, $\mathrm{scl}(g) = |\mathrm{rot}(g)|/2$. As observed above, many of these numbers are transcendental.

5.2.2. Bounded generation. For many specific (mostly nonuniform) lattices, the conclusion of Theorem 5.26 can be obtained directly by quite different methods.

DEFINITION 5.30. A group G is *boundedly generated* by a symmetric subset $H = H^{-1}$ if every element of G can be written as a product $h_1h_2\cdots h_n$ where each $h_i \in H$.

For this definition to be useful, the subset H should be small compared to G. The prototypical example of a boundedly generated group is $\mathrm{SL}(n,\mathbb{Z})$ where $n \ge 3$, or more generally $\mathrm{SL}(n, \mathcal{O})$ where $\mathcal{O}$ is the ring of integers of a number field (this fact is due to Carter–Keller). We do not state their theorem in full generality.

Definition 5.31. For $n \geq 3$ and $i \neq j \leq n$ the *elementary matrix* e_{ij} is the element of $\mathrm{SL}(n, \mathbb{Z})$ having 1's down the diagonal and in the ij location, and 0's elsewhere. An *elementary matrix* more generally is a power e_{ij}^m of some e_{ij}.

Theorem 5.32 (Carter–Keller [**53**]). *The group* $\mathrm{SL}(n, \mathbb{Z})$ *for* $n \geq 3$ *is boundedly generated by* elementary matrices. *In other words, there is a uniform bound* $N(n)$ *such that every element* $g \in \mathrm{SL}(n, \mathbb{Z})$ *can be written as a product of at most* N *elementary matrices.*

Example 5.33 ($\mathrm{SL}(n, \mathbb{Z})$ for $n \geq 3$). The stable commutator length vanishes identically on $\mathrm{SL}(n, \mathbb{Z})$ for $n \geq 3$. For, there is an identity

$$e_{ij}^n = [e_{ik}^n, e_{kj}]$$

provided i, j, k are distinct (which can be verified by direct calculation), and therefore $\mathrm{cl}(e_{ij}^n) = 1$ for all e_{ij} and all nonzero n. Since every $g \in \mathrm{SL}(n, \mathbb{Z})$ can be written as a product of a bounded number of powers of the e_{ij}, it follows that cl is uniformly bounded on $\mathrm{SL}(n, \mathbb{Z})$ and therefore scl vanishes identically.

In unpublished work, Carter–Keller and E. Paige extended these results considerably; Dave Witte-Morris [**159**] has obtained a very nice proof of their results using the Compactness Theorem of first-order logic. A special case of particular relevance is the following:

Theorem 5.34 (Carter–Keller–Paige [**159**] Thm. 6.1). *Let* A *be the ring of integers in a number field* K *(i.e. a finite algebraic extension of* $\mathbb{Q}$*) containing infinitely many units. Let* T *be an element of* $\mathrm{SL}(2, A)$ *which is not a scalar matrix (i.e. not of the form* $\lambda \cdot \mathrm{id}$*). Then* $\mathrm{SL}(2, A)$ *has a finite index normal subgroup which is boundedly generated by conjugates of* T*.*

Remark 5.35. If A is the ring of integers in a number field K, and A has only finitely many units, then K must be either $\mathbb{Q}$ or $\mathbb{Q}(\sqrt{-d})$ for some positive integer d. Every other A as above satisfies the hypothesis of the theorem.

Remark 5.36. The hypotheses of this theorem are equivalent to the property that $\mathrm{SL}(2, A)$ is isomorphic to an irreducible lattice in a higher rank semisimple Lie group. So the conclusion that scl vanishes identically also follows from Theorem 5.26.

Example 5.37. Let $A = \mathbb{Z}[\sqrt{2}]$, the ring obtained from $\mathbb{Z}$ by adjoining $\sqrt{2}$. Then $\Gamma = \mathrm{SL}(2, A)$ is boundedly generated by conjugates of $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Since $H_1(\mathrm{SL}(2, \mathbb{Z}); \mathbb{Z}) = \mathbb{Z}/6\mathbb{Z}$, the matrix T has a power which is a product of commutators in $\mathrm{SL}(2, \mathbb{Z})$, hence also in Γ. Let $H < \Gamma$ be a finite index normal subgroup of Γ which is boundedly generated by conjugates of T. Then cl is uniformly bounded on H, and therefore scl vanishes identically on H. Since H is finite index on Γ, every element of Γ has a power which is contained in H, hence scl vanishes identically on all of Γ.

The inclusion $\mathbb{Z}[\sqrt{2}] \to \mathbb{R}$ induces an inclusion of Γ into $\mathrm{SL}(2, \mathbb{R})$ whose image is dense. Let $\widehat{\Gamma}$ be the preimage in $\widetilde{\mathrm{SL}}(2, \mathbb{R})$. As in § 5.1 we conclude that $Q(\widehat{\Gamma})$ is one-dimensional, spanned by rotation number. Hence in $\widehat{\Gamma}$ we have $\mathrm{scl}(g) = |\mathrm{rot}(g)|/2$. By Gelfond–Schneider, these values are transcendental when they are not rational.

EXAMPLE 5.38. Another example, due to Liehl [**136**], says that $\mathrm{SL}(2,\mathbb{Z}[1/2])$ is boundedly generated by elementary matrices. As above, $Q(\Gamma)=0$, and $Q(\widehat{\Gamma})$ is one-dimensional, generated by rotation number, where Γ denotes $\mathrm{SL}(2,\mathbb{Z}[1/2])$ and $\widehat{\Gamma}$ its central extension. An element of Γ with trace 2^{-n} has transcendental rotation number when n is positive.

REMARK 5.39. If G is boundedly generated, so is a central extension $\widehat{G}$. Thus there are many examples of finitely presented groups which are boundedly generated, but for which $Q(\widehat{G})$ is nontrivial. This observation is made by Monod–Rémy in an appendix to [**143**]. They also observe that many of the groups G and $\widehat{G}$ furthermore have Kazhdan's property (T).

5.2.3. Symplectic groups. One class of Lie groups deserving special attention are the *symplectic groups*. As remarked earlier, there are two main sources of quasimorphisms. The first source, hyperbolic geometry, was studied systematically in Chapter 3. The second source is symplectic geometry (or more generally, causal or ordered structures); we turn to this subject in this section and the next. Basic references for symplectic geometry and topology are [**151**] and [**112**]. The material and exposition in this section borrows heavily from Barge–Ghys [**7**].

Given a vector space V (over $\mathbb{R}$ for simplicity), let V^* denote its dual. The *nth exterior product* $\Lambda^n V^*$, whose elements are called *n-forms on* V is the vector space generated by terms $v_1\wedge v_2\wedge\cdots\wedge v_n$ with the $v_i\in V^*$, which is linear in each factor separately, and subject to the relation that interchanging the order of two adjacent factors is multiplication by -1. With this notation, $\Lambda^1V^*=V^*$, and we make the convention that $\Lambda^0V^*=\mathbb{R}$. The sum $\oplus_i\Lambda^iV^*$ is a graded algebra, where multiplication is given by

$$v_1\wedge\cdots\wedge v_n\times u_1\wedge\cdots\wedge u_m=v_1\wedge\cdots\wedge v_n\wedge u_1\wedge\cdots\wedge u_m$$

and extended by linearity. If $x\in\Lambda^iV^*$ and $y\in\Lambda^jV^*$, then by counting signs, one sees that $xy=(-1)^{ij}yx$. If the dimension of V^* is m, then the dimension of Λ^iV^* is equal to $\binom{m}{i}$. Hence $\Lambda^mV^*\cong\mathbb{R}$, and $\Lambda^iV^*=0$ for all $i>m$.

DEFINITION 5.40. If V has dimension $2n$, a form $\omega\in\Lambda^2V^*$ is *symplectic* if $\omega\wedge\omega\wedge\cdots\wedge\omega\neq 0$ for any r-fold product, where $r\leq n$. Equivalently, $\omega^n\neq 0\in\Lambda^{2n}V^*$.

If G acts on V linearly, there is an induced action on V^* by the formula

$$g(v)(g(u))=v(u)$$

for all $v\in V^*$ and $u\in V$. This lets us define a diagonal action of G on each Λ^iV^* given by the formula

$$g(v_1\wedge\cdots\wedge v_n)=g(v_1)\wedge\cdots\wedge g(v_n)$$

and extended by linearity.

DEFINITION 5.41. Let V be a vector space and $\omega\in\Lambda^2V^*$ a symplectic form. The *symplectic group of* V,ω, denoted $\mathrm{Sp}(V,\omega)$, is the subgroup of $\mathrm{GL}(V)$ which fixes ω.

REMARK 5.42. When V has even dimension, the action of $\mathrm{GL}(V)$ on Λ^2V^* has a unique open dense orbit which consists exactly of the set of all symplectic elements of Λ^2V^*. It follows that any two groups $\mathrm{Sp}(V,\omega)$ and $\mathrm{Sp}(V,\omega')$ are conjugate as subgroups of $\mathrm{GL}(V)$, and their isomorphism class depends only on the dimension of V.

A vector space with an inner product may be identified with its dual. On $\mathbb{R}^{2n}$ with orthonormal basis $x_1, x_2, \cdots, x_n, y_1, y_2, \cdots, y_n$ there is a "standard" symplectic element given by the formula

$$\omega = x_1 \wedge y_1 + x_2 \wedge y_2 + \cdots + x_n \wedge y_n$$

Using the orthonormal basis to identify $\mathbb{R}^{2n}$ with its dual, this defines a symplectic form on $\mathbb{R}^{2n}$.

The symplectic group of $\mathbb{R}^{2n}$ with respect to ω is usually called *the symplectic group*, and denoted $\mathrm{Sp}(2n, \mathbb{R})$. If J denotes the $2n \times 2n$ matrix whose four $n \times n$ blocks have the form

$$J = \begin{pmatrix} 0 & \mathrm{id} \\ -\mathrm{id} & 0 \end{pmatrix}$$

then $\mathrm{Sp}(2n, \mathbb{R})$ is the group of matrices A for which $A^T J A = J$.

Let $\mathrm{U}(n)$ denote the *unitary group*, i.e. the group of $n \times n$ complex matrices which preserve the standard Hermitian inner product on $\mathbb{C}^n$. If we think of $\mathbb{R}^{2n}$ as the underlying real vector space of $\mathbb{C}^n$, then the inclusion $M_n(\mathbb{C}) \to M_{2n}(\mathbb{R})$ realizes $\mathrm{U}(n)$ as a compact subgroup of $\mathrm{Sp}(2n, \mathbb{R})$. In fact, $\mathrm{U}(n)$ is a *maximal* compact subgroup, and the coset space $X := \mathrm{Sp}(2n, \mathbb{R})/\mathrm{U}(n)$ admits an $\mathrm{Sp}(2n, \mathbb{R})$-invariant Riemannian metric of non-positive curvature. The space X is usually called the *Siegel upper half-space*, and has several equivalent descriptions. One well-known description says that X is the space of $n \times n$ complex symmetric matrices whose imaginary part is positive definite. If $n = 1$, this is the set of complex numbers with positive imaginary part, which is the upper half-space model of the (ordinary) hyperbolic plane.

Since X is non-positively curved and complete, it is contractible, so the inclusion $\mathrm{U}(n) \to \mathrm{Sp}(2n, \mathbb{R})$ is a homotopy equivalence. The group $\mathrm{U}(n)$ acts transitively on the unit sphere S^{2n-1} in $\mathbb{C}^n$, with stabilizer $\mathrm{U}(n-1)$, so there is a fibration

$$\mathrm{U}(n-1) \to \mathrm{U}(n) \to S^{2n-1}$$

By the homotopy exact sequence of a fibration, it follows that $\pi_1(\mathrm{U}(n)) = \mathbb{Z}$, generated by the inclusion $S^1 = \mathrm{U}(1) \to \mathrm{U}(n)$, and therefore $\pi_1(\mathrm{Sp}(2n, \mathbb{R})) = \mathbb{Z}$.

Let $\widetilde{\mathrm{Sp}}(2n, \mathbb{R})$ denote the universal covering group. In the case $n = 1$ this is just $\widetilde{\mathrm{SL}}(2, \mathbb{R})$.

A closed differential 2-form ω on a manifold M^{2n} of dimension $2n$ is *symplectic* if the $2n$-form ω^n is nonzero at every point. It turns out that there is a natural symplectic form ω on the Siegel upper half-space X which is invariant under $\mathrm{Sp}(2n, \mathbb{R})$. If Γ is a (torsion-free) lattice in $\mathrm{Sp}(2n, \mathbb{R})$, then ω descends to a symplectic form on X/Γ. If Γ is cocompact, the cohomology class $[\omega] \in H^2(X/\Gamma) = H^2(\Gamma)$ is nonzero, since the integral of the top power of ω over X/Γ is nonzero. In fact, it turns out that the class of $[\omega]$ is in the image of $H^2_b(\Gamma)$. Moreover, Domic and Toledo [**66**] calculate the norm of this class, and show that it is equal to $n\pi$.

If we let $\widehat{\Gamma}$ denote the preimage of Γ in $\widetilde{\mathrm{Sp}}(2n, \mathbb{R})$, then $[\omega]$ pulls back to a class $[\tilde{\omega}]$ in $H^2_b(\widehat{\Gamma})$ whose image in $H^2(\widehat{\Gamma})$ is trivial, and therefore comes from a homogeneous quasimorphism ρ, which we normalize by scaling to have $D(\rho) = n$. Evidently, in the case $n = 1$, the quasimorphism ρ is just rotation number. Barge–Ghys [**7**] call this quasimorphism the *symplectic rotation number*. In fact, since the form ω is invariant under the action of $\mathrm{Sp}(2n, \mathbb{R})$ on X, there is a well-defined homogeneous quasimorphism ρ defined on the entire group $\widetilde{\mathrm{Sp}}(2n,\mathbb{R})$, and $\rho \in Q(\widehat{\Gamma})$ is just pulled back by inclusion.

Barge–Ghys give an explicit description for this quasimorphism, as follows.

DEFINITION 5.43. A subspace π of $\mathbb{R}^{2n}$ of real dimension n is *Lagrangian* if the symplectic form ω restricts to zero on π. That is, if $\omega(u,v) = 0$ for all $u, v \in \pi$.

A subspace π of $\mathbb{R}^{2n}$ of real dimension n is Lagrangian if and only if it is totally real when considered as a subspace of $\mathbb{C}^n$. It follows that the subgroup $\mathrm{U}(n)$ of $\mathrm{Sp}(2n,\mathbb{R})$ acts transitively on the space Λ_n of Lagrangian subspaces of $\mathbb{R}^{2n}$, with stabilizer the subgroup $\mathrm{O}(n,\mathbb{R})$. In other words, there is an isomorphism $\mathrm{U}(n)/\mathrm{O}(n,\mathbb{R}) = \Lambda_n$ as principal $\mathrm{U}(n)$-spaces. Note that we are thinking here of $\mathrm{O}(n,\mathbb{R})$ firstly as a subgroup of $\mathrm{GL}(n,\mathbb{C})$ by the inclusion $\mathbb{R} \to \mathbb{C}$, and then secondly as a subgroup of $\mathrm{Sp}(2n,\mathbb{R})$ by the inclusion $\mathrm{GL}(n,\mathbb{C}) \to \mathrm{GL}(2n,\mathbb{R})$ coming from the identification of $\mathbb{C}^n$ with $\mathbb{R}^{2n}$.

Let $* \in \Lambda_n$ be some basepoint, for example corresponding to the Lagrangian subspace $\mathbb{R}^n$ in $\mathbb{C}^n$. For each $g \in \mathrm{Sp}(2n,\mathbb{R})$, there is a unique coset $u(g)\mathrm{O}(n,\mathbb{R}) \in \mathrm{U}(n)/\mathrm{O}(n,\mathbb{R})$ with $g(*) = u(g)(*)$ for any element of the coset. The homomorphism $\det^2 : \mathrm{U}(n) \to S^1$ factors through the quotient $\mathrm{U}(n)/\mathrm{O}(n,\mathbb{R})$, and defines a function $\det^2 : \mathrm{Sp}(2n,\mathbb{R}) \to S^1$. This map is a double covering, restricted to the subgroup $S^1 = \mathrm{U}(1)$, so we get a covering map

$$\mu : \widetilde{\mathrm{Sp}}(2n,\mathbb{R}) \to \mathbb{R}$$

which turns out to be a quasimorphism.

The quasimorphism μ is not homogeneous. However, Barge–Ghys derive a formula for its homogenization ρ, at least mod $\mathbb{Z}$. To state their theorem we must first recall some standard facts about the spectrum of a symplectic matrix. Let $A \in \mathrm{Sp}(2n,\mathbb{R})$ and suppose for simplicity that A is diagonalizable over $\mathbb{C}$. The spectrum of A (i.e. the set of complex eigenvalues with multiplicity) is invariant under conjugation, since A is a real matrix. Moreover, it is invariant with respect to inversion in the unit circle in $\mathbb{C}$. Hence if λ is an eigenvalue, then $\lambda, \lambda^{-1}, \overline{\lambda}, \overline{\lambda}^{-1}$ are all eigenvalues. The case that λ is real or on the unit circle is naturally rather special. It turns out that eigenvalues λ which are not on the unit circle do not contribute to ρ.

Suppose A is diagonalizable over $\mathbb{C}$, and H is the subspace of $\mathbb{R}^{2n}$ of dimension $2k$ spanned by the 2×2 Jordan blocks of A (over $\mathbb{R}$) corresponding to pairs of complex eigenvalues $\lambda, \overline{\lambda}$ with λ on the unit circle. Then H is a symplectic subspace of $\mathbb{R}^{2n}$, and the restriction of A to H is orthogonal, and therefore unitary; hence $A|_H$ is conjugate in the symplectic group to a unitary matrix $B \in \mathrm{U}(k)$. The complex eigenvalues of B are called the *proper values* of A of absolute value 1.

Barge–Ghys' theorem gives a formula for ρ in terms of the proper values of absolute value 1.

THEOREM 5.44 (Barge–Ghys [**7**], Thm. 2.10). *Let g be an element of* $\mathrm{Sp}(2n,\mathbb{R})$*, and let $\lambda_1, \cdots, \lambda_k$ be the proper values of g of absolute value* 1*, listed with multiplicity. Then*

$$\rho(g) = \frac{1}{\pi}\sum \arg(\lambda_i) \pmod{\mathbb{Z}}$$

REMARK 5.45. If we deform a matrix in $\mathrm{Sp}(2n,\mathbb{R})$ so that some set $\{\lambda, \lambda^{-1}, \overline{\lambda}, \overline{\lambda}^{-1}\}$ of eigenvalues is deformed onto the unit circle, one obtains for the deformed matrix two proper values of absolute value one, which are equal to λ and $\overline{\lambda}$ respectively, and therefore the sum of their arguments vanishes. This explains why ρ is continuous on $\mathrm{Sp}(2n,\mathbb{R})$, which is otherwise not obvious.

Remark 5.46. If we think of $\mathbb{R}^{2n}$ with its standard symplectic form as a product of n copies of $\mathbb{R}^2$ with its standard symplectic form, we get a natural inclusion

$$\mathrm{SL}(2,\mathbb{R}) \times \mathrm{SL}(2,\mathbb{R}) \times \cdots \times \mathrm{SL}(2,\mathbb{R}) \to \mathrm{Sp}(2n,\mathbb{R})$$

The symplectic rotation number restricts to Poincaré's rotation number on each $\mathrm{SL}(2,\mathbb{R})$ factor, and is equal to the sum of rotation numbers on the factors on the image of the product of $\mathrm{SL}(2,\mathbb{R})$'s.

Theorem 5.26 shows that $H_b^2(\Gamma)$ includes into $H^2(\Gamma)$, when n is at least 2. Since the defect of ρ is n, there is a formula

$$\mathrm{scl}(g) = |\rho(g)|/2n = \frac{1}{2n\pi}\sum \arg(\lambda_i) \pmod{\frac{1}{2n}\mathbb{Z}}$$

for $g \in \widehat{\Gamma}$. Lattices in $\mathrm{Sp}(2n,\mathbb{R})$ for n at least 2 have algebraic entries. Hence by Gelfond–Schneider, scl is transcendental on $\widehat{\Gamma}$ when it is irrational.

Obviously the examples above can be generalized tremendously. However in every case, the irrational values of scl obtained appear to be transcendental. Hence we pose the following question.

Question 5.47. *Is there a finitely presented group G in which* scl *takes on an irrational value that is algebraic?*

More generally, one can ask for a complete characterization of the values of scl that can occur in finitely presented groups.

Question 5.48. *What real numbers are values of* scl *on elements in finitely presented groups?*

This seems like a difficult question.

5.2.4. Causal structures and quasimorphisms. In this section we give a more topological definition of the symplectic rotation quasimorphism ρ defined in § 5.2.3 which "explains" the integral value of $D(\rho)$. The construction makes use of the *causal* structure on Λ_n. This point of view is particularly explicit in [**3**]. Also compare [**54**].

Definition 5.49. Let V be a real vector space. A *cone* C in V is a subset of the form $\mathbb{R}\cdot K$ where K is compact and convex with nonempty interior, and disjoint from the origin. A vector $v \in V$ is *timelike* if it is in the interior of C, is *lightlike* if it is in the frontier of C, and is *spacelike* otherwise.

Example 5.50. Let V be an $(n+1)$-dimensional real vector space, and $q : V \times V \to \mathbb{R}$ a symmetric bilinear pairing of signature $(n,1)$ (i.e. with n positive eigenvalues and one negative eigenvalue). The set of vectors v with $q(v,v) \leq 0$ is a cone in V.

If M is a smooth manifold, a *cone field* is a continuously varying choice of cone in the tangent space at each point. The set of timelike vectors at a point has two components; a *causal structure* on M is a cone field together with a continuously varying choice of one of these components (the *positive* cone) at each point. Two points p, q are *causally connected*, and we write $p \prec q$, if there is a nontrivial smooth curve from p to q whose tangent vector at every point is positive and timelike. The relation $\prec$ is transitive (but *not* typically reflexive or symmetric). A causal structure is *recurrent* if $p \prec q$ for all p and q.

REMARK 5.51. Some authors use the notational convention that $p \prec q$ means either that $p = q$ or that p is causally connected to q in the sense above. We denote this instead by $p \preceq q$.

Let M be a closed manifold which admits a recurrent causal structure, and let S be a non-separating codimension one submanifold whose tangent space is spacelike. Then S is essential in homology, and is dual to an element of $H^1(M;\mathbb{Z})$. Let M' denote the infinite cyclic cover of M dual to S. The causal structure on M lifts to one on M' (where it is no longer recurrent).

Let $C^+(M)$ denote the group of diffeomorphisms of M which preserve the causal structure, and $C^+(M')^{\mathbb{Z}}$ the preimage of this group in $\text{Homeo}^+(M')$. There is a central extension

$$0 \to \mathbb{Z} \to C^+(M')^{\mathbb{Z}} \to C^+(M) \to 0$$

where $\mathbb{Z}$ is the deck group. We write the action of the deck group on points in M' by $p \to p + n$.

For any $p, q \in M'$, define $d(p,q)$ to be the greatest integer $n \in \mathbb{Z}$ such that $p \prec q - n$. Pick a basepoint $*$ in M', and for any $\alpha \in C^+(M')^{\mathbb{Z}}$, define $\phi(\alpha) = d(*, \alpha(*))$ and $\rho(\alpha) = \lim_{n\to\infty} \phi(\alpha^n)/n$. Since the causal structure on M is recurrent, there is a least positive integer w such that any two points p and q are contained in a closed timelike curve which intersects S at most w times.

LEMMA 5.52. *The function ϕ as above is a quasimorphism, and ρ is its homogenization. Moreover, the defect of ρ is at most w.*

PROOF. For any α there is equality $\phi(\alpha - \phi(\alpha)) = 0$. Let α, β be arbitrary, and denote $\alpha' = \alpha - \phi(\alpha)$ and $\beta' = \beta - \phi(\beta)$. Then $* \prec \alpha'(*) \prec * + w$ and similarly for $\beta'(*)$. We calculate

$$* \prec \alpha'(*) \prec \alpha'\beta'(*) \prec \alpha'(* + w) \prec * + 2w$$

and therefore

$$|\phi(\alpha\beta) - \phi(\alpha) - \phi(\beta)| = |\phi(\alpha'\beta')| \leq 2w$$

This shows that ϕ is a quasimorphism; evidently ρ is its homogenization.

To estimate the defect of ρ we repeat the argument of Lemma 2.41. For any $p \in M'$ and any elements $\alpha, \beta \in C^+(M')^{\mathbb{Z}}$, after multiplying by elements of the center if necessary, we can assume

$$p \preceq \alpha(p) \preceq \alpha\beta(p) \prec \alpha(p + w) \prec p + 2w$$

$$p \preceq \beta(p) \preceq \beta\alpha(p) \prec \beta(p + w) \prec p + 2w$$

Set $q = \beta\alpha(p)$. Then $p \preceq q \prec p + 2w$ and therefore

$$q - 2w \prec p \preceq \alpha\beta(p) = [\alpha, \beta](q) \prec p + 2w \preceq q + 2w$$

Since p was arbitrary, so was q, and we have shown that $q - 2w \prec [\alpha, \beta](q) \prec q + 2w$ for any q and any commutator $[\alpha, \beta]$.

It follows that if γ is a product of m commutators, then $|\rho(\gamma)| \leq 2w(m + 1)$. Taking m large, the argument of Lemma 2.24 shows $D(\rho) \leq w$. □

REMARK 5.53. Essentially the same construction is described in [**54**], § 7–8.

Causal structures arise naturally in certain contexts.

EXAMPLE 5.54. Let $\mathfrak{G}$ be a simple Lie algebra with Lie group G. An $\mathrm{Ad}(G)$-invariant cone in $\mathfrak{G}$ exponentiates to a G-invariant cone field on G. This determines a causal structure either on G or on a double cover, which is invariant under the action of the group on itself. Let K be a maximal compact subgroup of G, with Lie algebra $\mathfrak{k}$. It turns out (Paneitz [**165**], Cor. 3.2) that there is an $\mathrm{Ad}(G)$-invariant cone in $\mathfrak{G}$ if and only if $\mathfrak{k}$ has nontrivial center.

If $G = \mathrm{Sp}(2n,\mathbb{R})$, then

$$\mathfrak{G} = \begin{pmatrix} A & B \\ C & -A^t \end{pmatrix} \text{ where } A,B,C \text{ are } n \times n \text{ blocks, and } B,C \text{ are symmetric}$$

$$\mathfrak{k} = \begin{pmatrix} A & -B \\ B & A \end{pmatrix} \text{ where } A \text{ is skew, and } B \text{ is symmetric}$$

The center of $\mathfrak{k}$ is nontrivial, and spanned by the matrix $\left(\begin{smallmatrix} 0 & -\mathrm{Id} \\ \mathrm{Id} & 0 \end{smallmatrix}\right)$. If ω denotes the standard (and $\mathrm{Ad}(G)$-invariant) symplectic form on $\mathbb{R}^{2n}$, define C to be the cone of vectors $X \in \mathfrak{G}$ for which $\omega(\mathrm{ad}(X)v, v) \ge 0$ for all $v \in \mathbb{R}^{2n}$. This is nonempty and invariant, and defines a (recurrent) causal structure on $\mathrm{Sp}(2n,\mathbb{R})$.

EXAMPLE 5.55. Let $G = \mathrm{SO}(n,2)$, the group of linear automorphisms of $\mathbb{R}^{n+2}$ which preserve the quadratic form $q(x) = x_1^2 + \cdots + x_n^2 - x_{n+1}^2 - x_{n+2}^2$. Let H be the hyperboloid of vectors x for which $q(x) = -1$. Then G acts transitively on H. At a point $x \in H$, the tangent space T_xH is naturally isomorphic to the orthogonal subspace of $\mathbb{R}^{n+2}$ to x with respect to the form q. Since $q(x) = -1$, the restriction of q to this subspace has signature $(n,1)$, and therefore G preserves a cone field on H as in Example 5.50. There is a subgroup $\mathrm{SO}_0(n,2)$ of index 2 which preserves the orientation on the cone field, and therefore a causal structure on H.

When $n = 1$, the group $\mathrm{SO}(1,2)$ is isomorphic to $\mathrm{PSL}(2,\mathbb{R})$, the group of isometries of the hyperbolic plane. In the Klein (projective) model, the hyperbolic plane is identified with the interior of a round disk D in $\mathbb{RP}^2$, and the exterior $\mathbb{RP}^2 - D$ (which is homeomorphic to an open Möbius strip) is equal to $H/\pm 1$ where H is as in Example 5.55. If p is a point in $\mathbb{RP}^2 - D$, there are two straight lines through p which are tangent to ∂D. The cone at p is the set of tangents to straight lines through p which do not intersect D. A smooth curve in $\mathbb{RP}^2 - D$ is timelike if every tangent line to the curve is disjoint from D. Evidently, the causal structure on H is recurrent; in fact, one sees that any two points in H are contained in a closed timelike loop with winding number at most 2. By rotational symmetry, it follows that the same is true for arbitrary $n \ge 2$ and therefore one obtains a homogeneous quasimorphism on the universal covering group $\widetilde{\mathrm{SO}}_0(n,2)$ with defect at most 2. When $n \ge 2$, this estimate can be seen to be sharp by an explicit construction (compare with Domic–Toledo [**66**] and [**55**]).

Causal structures on noncompact manifolds often extend to causal structures on certain natural boundaries. A *symmetric bounded domain* is a complex symmetric space that is isomorphic to a bounded domain in $\mathbb{C}^n$ for some n. It is *irreducible* if its universal cover is not a nontrivial direct product of symmetric spaces. By a theorem of Harish-Chandra, every irreducible complex symmetric space of noncompact type is bounded. An irreducible symmetric bounded domain is said to be of *tube type* if it is isomorphic to a domain of the form $V + i\Omega$ where $\Omega \subset V$ is a proper open cone in the real vector space V.

A realization of a bounded symmetric domain defines a natural compactification. The group G of holomorphic automorphisms of the domain extends to the compactification, and the *Shilov boundary* is the unique closed G-orbit in the compactification. It is known (see e.g. [**120**], § 5) that the Shilov boundary of a symmetric bounded domain of tube type admits a natural causal structure.

EXAMPLE 5.56. The Siegel upper half-space $\mathrm{Sp}(2n,\mathbb{R})/\mathrm{U}(n)$ is a symmetric bounded domain of tube type. Its Shilov boundary is the space Λ_n of Lagrangians in $\mathbb{R}^{2n}$.

The causal structure on Λ_n can be given a very geometric definition, as observed by Arnold [**3**]. If π is a Lagrangian subspace of $\mathbb{R}^{2n}$ (and therefore corresponds to a point in Λ_n) the *train* of π is the set of Lagrangian subspaces of $\mathbb{R}^{2n}$ which are not transverse to π.

Fix a Lagrangian π and a transverse Lagrangian σ, and let π_t be a 1-parameter family of Lagrangians with $\pi_0 = \pi$. For small t, the Lagrangians π_t and σ are still transverse, and span $\mathbb{R}^{2n}$. For each $v \in \mathbb{R}^{2n}$ and each such t, there is a unique decomposition $v = v(\pi_t) + v(\sigma)$ where $v(\pi_t) \in \pi_t$ and $v(\sigma) \in \sigma$ (note that for a fixed v, the vector $v(\sigma)$ typically depends on t). Define a 1-parameter family of bilinear forms q_t on $\mathbb{R}^{2n}$ by the formula

$$q_t(v,w) = \omega(v(\pi_t), w(\sigma))$$

where ω is the symplectic form. In this way, a tangent vector $\pi_0' := \frac{d}{dt}\big|_0 \pi_t$ to π determines a *symmetric* bilinear form $q_0' := \frac{d}{dt}\big|_0 q_t$ which vanishes identically on σ, and can be thought of as a symmetric bilinear form on π. The map $\pi_0' \to q_0'$ is an isomorphism from the tangent space $T_\pi\Lambda_n$ to the space of symmetric bilinear forms on π (to see this, observe that it is linear and injective, and is surjective by a dimension count, since both $\mathrm{U}(n)/\mathrm{O}(n)$ and the space of symmetric $n \times n$ matrices have dimension $n(n+1)/2$). Note that q_0' is degenerate precisely along the subspace $\pi_0' \cap \pi$. Hence the tangent cone to the train at π corresponds precisely to the degenerate bilinear forms. Exponentiating, we see that in a neighborhood of π, the train separates Λ_n into chambers, corresponding to nondegenerate quadratic forms on $\mathbb{R}^n$ of a fixed signature. The *positive cone* corresponds (infinitesimally) to positive definite quadratic forms on π.

EXAMPLE 5.57. The space $\Lambda_2 = \mathrm{U}(2)/\mathrm{O}(2)$ is diffeomorphic to the nonorientable sphere bundle over S^1. Fix co-ordinates $\Lambda_2 = S^2 \times [0,1]/\sim$ where $(\theta,0) \sim (-\theta,1)$. Fix a basepoint $*$ to be the north pole of the sphere $S^2 \times 0$ in these co-ordinates. The train of $*$ intersects each sphere $S^2 \times t$ in a circle of constant latitude which decreases monotonically with t, until it converges to the south pole in $S^2 \times 1$ (which is identified with $*$ by the holonomy map).

EXAMPLE 5.58. Let $G = \mathrm{SO}(n,2)$, and recall the notation from Example 5.55. The projectivization of the cone $q = 0$ is an S^{n-1} bundle over S^1 that we denote by E (this bundle is twisted by the antipodal map, so E is topologically a product if and only if n is even). Then E is a Shilov boundary for G. In the projectivization, E divides $\mathbb{RP}^{n+1}$ into two components, one of which is $H/\pm 1$. The cone field on H limits to a cone field on E, where the cone at a point $e \in E$ is the set of tangent lines to E which point into $H/\pm 1$. The group $\mathrm{SO}_0(n,2)$ preserves the causal structure on E. When $n = 1$, E is a circle, which can be thought of as the circle at infinity of the hyperbolic plane. When $n = 2$, E is a torus, and the cone structure determines a pair of transverse foliations on this torus by circles. $\mathrm{SO}_0(2,2)$ acts on the leaf spaces of these foliations (which are themselves circles) by projective transformations, exhibiting the exceptional 2-fold covering $\mathrm{SO}_0(2,2) \to \mathrm{PSL}(2,\mathbb{R}) \times \mathrm{PSL}(2,\mathbb{R})$. When $n = 3$, E is a twisted S^2 bundle over S^1, and is equal to the space Λ_2 as described in Example 5.57; this reflects the exceptional isomorphism $\mathrm{SO}_0(3,2) = \mathrm{Sp}(4,\mathbb{R})/\pm 1$.

Causal structures become very rigid in high (≥ 3) (real) dimensions. For example, one has the following:

THEOREM 5.59 (Kaneyuki [**120**], Thm. 6.2). *Let D be an irreducible symmetric bounded domain of tube type, and $G(D)$ the group of holomorphic automorphisms of D. Let S be the Shilov boundary of D with its natural causal structure. Let $C^+(S)$ be the group of causal homeomorphisms of S. Suppose (complex)* $\dim(D) > 1$. *Then $C^+(S) = G(D)$.*

5.3. Braid groups and transformation groups

5.3.1. Braid groups.

DEFINITION 5.60. The *braid group B_n* on n strands is generated by elements σ_i for $i = 1, 2, \cdots, n-1$ and relations $[\sigma_i, \sigma_j] = 1$ when $|i - j| \neq 1$, and $\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}$.

These groups were introduced by Emil Artin in 1925 [**5**].

A word in the generators is represented pictorially by a projection of a tangle of n arcs running between two parallel vertical lines, where no arc has any vertical tangencies. Braids are composed by "gluing" pictures; see Figure 5.2. A generator

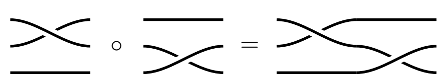


FIGURE 5.2. Braids are represented by pictures; composition is performed by gluing adjacent pictures. This picture illustrates the composition of σ_1 with σ_2^{-1} in B_3.

σ_i is represented by a crossing, where the ith strand crosses over the $(i+1)$st strand, and σ_i^{-1} is represented by a crossing where the $(i+1)$st strand crosses over the ith strand. Equivalence in B_n corresponds to equivalence of pictures up to "isotopy". The relation $[\sigma_i, \sigma_j] = 1$ when $|i - j| \neq 1$ corresponds to the fact that crossings on disjoint pairs of strands can be performed in either order. The group law $\sigma_i^{-1}\sigma_i = \sigma_i \sigma_i^{-1} = \mathrm{id}$ corresponds to the Reidemeister 2 move on diagrams, and the relation $\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}$ corresponds to the Reidemeister 3 move on diagrams; see Figure 5.3.

Another way to think of B_n is as a mapping class group. A diagram of a braid can be thought of as a tangle in a product $D^2 \times [0,1]$ transverse to the foliation by vertical disks. In this way, an element in B_n determines a loop in the configuration space of distinct n-tuples of points in the disk. Isotopy of braids corresponds to

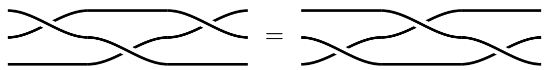


FIGURE 5.3. The relation $\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}$ in the braid group corresponds to the Reidemeister 3 move on diagrams.

homotopy of loops, so B_n can be thought of as the fundamental group of the space of distinct n-tuples in D^2. Equivalently, B_n is just the mapping class group rel. boundary of a disk with n punctures. Braid groups, as examples of mapping class groups, admit a very large space of homogeneous quasimorphisms, by the construction described in § 3.5.

Gambaudo–Ghys [**87**] use symplectic geometry to define some quite different quasimorphisms. Many interesting representations of B_n can be derived from their geometric description as mapping class groups. Let D_n denote the disk with n points removed. There is an isomorphism $\pi_1(D_n) \to F_n$, the free group on n generators, and the generators may be taken to be loops, each of which winds around one puncture. Let $\epsilon : \pi_1(D_n) \to \mathbb{Z}$ take each generator to 1. This homomorphism defines a cyclic cover $\widetilde{D}_n$, whose first homology $H_1(\widetilde{D}_n)$ can be thought of as a $\mathbb{Z}[q, q^{-1}]$-module, where q generates the deck group of the covering. The first homology group is free as a module of rank $(n-1)$. If e_i is a based loop in D_n winding positively once around the ith puncture, the loops $\alpha_i := e_{i+1}e_i^{-1}$ for $1 \le i \le (n-1)$ all lift to $\widetilde{D}_n$, and freely generated $H_1(\widetilde{D}_n)$ as a $\mathbb{Z}[q, q^{-1}]$-module.

If we fix some basepoint $p \in D_n$, every braid $\psi \in B_n$ is represented by a homeomorphism which fixes p, and is covered by a unique homeomorphism $\widetilde{\psi}$ of $\widetilde{D}_n$ which fixes the preimages of p pointwise. Hence there is an induced action of B_n on $H_1(\widetilde{D}_n)$ by $\mathbb{Z}[q, q^{-1}]$-module automorphisms, and thereby a representation $\beta : B_n \to \mathrm{GL}(n-1, \mathbb{Z}[q, q^{-1}])$. This representation is called the *Burau representation*. See e.g. [**15**] for an elegant geometric interpretation of this action, and [**16**] as a general reference for braid groups. As matrices, this representation has the form

$$\sigma_1 \to \begin{pmatrix} -q^{-1} & q^{-1} \\ 0 & 1 \end{pmatrix} \oplus \mathrm{Id}_{n-3}, \quad \sigma_{n-1} \to \mathrm{Id}_{n-3} \oplus \begin{pmatrix} 1 & 0 \\ 1 & -q^{-1} \end{pmatrix},$$

and

$$\sigma_i \to \mathrm{Id}_{i-2} \oplus \begin{pmatrix} 1 & 0 & 0 \\ 1 & -q^{-1} & q^{-1} \\ 0 & 0 & 1 \end{pmatrix} \oplus \mathrm{Id}_{n-i-2} \text{ for } 1 < i < n-1$$

where the notation $A \oplus B$ stands for the block matrix $\left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right)$.

REMARK 5.61. Several different conventions exist in the literature, depending on whether one takes σ_i or σ_i^{-1} as the generators of B_n, and whether one studies the action on homology or cohomology.

Squier [**190**] showed that the image of the Burau representation is unitary, in the following sense. It turns out that there is a nonsingular matrix J_0 defined over $\mathbb{Z}[q, q^{-1}]$ such that for each $w \in B_n$, one has $\beta(w)^* J_0 \beta(w) = J_0$ (here $*$ is the conjugate transpose, where conjugation interchanges q with q^{-1}). In fact, over $\mathbb{Z}[s, s^{-1}]$ where $s^2 = q$, a change of basis replaces J_0 by a matrix J satisfying $J^* = J$.

If $\sigma : \mathbb{Z}[s, s^{-1}] \to \mathbb{C}$ takes s to an element of norm 1, the matrix $J(s)$ is Hermitian (in the usual sense) and one obtains a representation $\beta_\sigma : B_n \to \mathrm{U}(J)$, the unitary group of the form J. If J is nondegenerate, its imaginary part is a nondegenerate antisymmetric form, and one therefore obtains a representation $\beta_\sigma : B_n \to \mathrm{Sp}(2n-2; \mathbb{R})$. It turns out that the forms J are degenerate exactly when s is a $(2n)$th root of unity different from ± 1 (so that q is an nth root of unity different from 1). When s is sufficiently close to 1, the form J is positive definite. Each time q crosses an nth root of unity, the number of positive eigenvalues changes by -1. So when q is specialized to an mth root of unity with $m < n$ and m, n coprime, the form is nondegenerate, the signature is indefinite, and the image of β_σ in $\mathrm{Sp}(2n-2; \mathbb{R})$ typically has noncompact closure.

Another way to obtain these representations of B_n is by using surface topology. For each m, let $D_{n,m}$ be the surface obtained by taking an m-fold branched cover of the disk over n points. The induced action of B_n on $D_{n,m}$ is well-defined up to homotopy, and we get a representation on the vector space $H_1(D_{n,m}, \partial D_{n,m}; \mathbb{R})$. The deck group $\mathbb{Z}/m\mathbb{Z}$ acts on $D_{n,m}$. If ω is an mth root of unity, the ω-eigenspace of this action is real, and B_n-invariant. There is thus an action of B_n on the invariant vector space $H_1(D_{n,m}, \partial D_{n,m}; \mathbb{R})_\omega$. It turns out this representation is isomorphic to the Burau representation evaluated at $q = \omega$ (see e.g. [**87**], Prop. 2.2). The ordinary intersection pairing on H_1 is nondegenerate on this subspace when n and m are coprime, and one sees in another way the symplectic structure.

REMARK 5.62. When n and m are not coprime, the imaginary part of J is degenerate on a subspace, and one obtains a symplectic action of B_n on the quotient by this subspace.

The cohomology of classical braid groups was computed by Arnold [**1**] (also see [**201**], Thm. 4.1). He showed the following:

THEOREM 5.63 (Arnold [**1**]). *For $n \geq 2$, there are isomorphisms $H^0(B_n; \mathbb{Z}) = H^1(B_n; \mathbb{Z}) = \mathbb{Z}$. Otherwise, $H^i(B_n; \mathbb{Z})$ is finite when $i \geq 2$ and zero when $i \geq n$.*

We are concerned with the case $i = 2$. Theorem 5.63 says that $H^2(B_n; \mathbb{Z})$ is torsion. Consequently, each representation $\beta_\sigma : B_n \to \mathrm{Sp}(2n-2, \mathbb{R})$ defines a quasimorphism ρ on B_n (well-defined up to elements of H^1), whose coboundary is the pullback of the generator of $H^2_b(\mathrm{Sp}(2n-2))$ under β^*_σ.

EXAMPLE 5.64. The braid group B_3 is discussed in Example 4.33. In the special case of B_3, the image of the Burau representation evaluated at -1 is equal to $\mathrm{SL}(2, \mathbb{Z})$, and ρ is the rotation quasimorphism coming from the action of $\mathrm{PSL}(2, \mathbb{Z})$ on S^1. A slightly different normalization of this quasimorphism is sometimes called the *Rademacher function* on $\mathrm{SL}(2, \mathbb{Z})$; see § 4 of [**87**], and § 6.1.7.

EXAMPLE 5.65. The Burau representation of B_4 evaluated at $\omega = e^{2\pi i/3}$ is 3 (complex) dimensional, and has matrix entries in the discrete subring $\mathbb{Z}[\omega]$ of $\mathbb{C}$. The form J has signature $(1, 2)$. Projectivizing, one obtains a discrete representation of B_4 into $\mathrm{PU}(1, 2)$, the group of isometries of the complex hyperbolic plane. One may therefore obtain interesting de Rham quasimorphisms on B_4, as in § 2.3.1.

5.3.2. Area-preserving diffeomorphisms of surfaces. Gambaudo–Ghys [**86**] showed how to use quasimorphisms on discrete groups to obtain nontrivial quasimorphisms on certain transformation groups.

Similar ideas appeared earlier in work of Arnold [**4**], Ruelle [**181**], Gambaudo–Sullivan–Tresser [**88**] and others. A given (continuous) dynamical system is approximated (in some sense) by a discrete combinatorial model. Associated to the discrete approximation is some numerical invariant, which can then be integrated over the degrees of freedom of the continuous system. For this integration to make sense and have useful properties, the continuous dynamical system must be (at least) measure preserving, and of sufficient regularity that the integral converges.

The case presenting the fewest technical details is that of a group of area-preserving diffeomorphisms of a (finite area) surface.

DEFINITION 5.66. For any surface S, let $\text{Diff}^\infty(S, \partial S, \text{area})$ (or omit the ∂S in the notation if S has no boundary) denote the group of diffeomorphisms of S, fixed pointwise on the boundary, that preserve the (standard) area form, and let $\text{Diff}_0^\infty(S, \partial S, \text{area})$ denote the subgroup of such diffeomorphisms isotopic to the identity.

There is an exact sequence

$$\text{Diff}_0^\infty(S, \partial S, \text{area}) \to \text{Diff}^\infty(S, \partial S, \text{area}) \to \text{MCG}(S, \partial S)$$

Quasimorphisms on mapping class groups can be pulled back to $\text{Diff}^\infty(S, \partial S, \text{area})$. Therefore we focus on the construction of quasimorphisms on $\text{Diff}_0^\infty(S, \partial S, \text{area})$. A key case to consider is $S = D$, the closed unit disk.

DEFINITION 5.67. Fix some n, and let μ be a quasimorphism on B_n. Fix n distinct points x_i^0 in D for $1 \le i \le n$. Given $g \in \text{Diff}_0^\infty(D, \partial D, \text{area})$, let g_t be an isotopy from id to g. For a generic ordered n-tuple of distinct points $x_1, \cdots, x_n$ in D, let $\gamma(g; x_1, \cdots, x_n) \in B_n$ be the braid obtained by first moving the x_i^0 in a straight line to the x_i, then composing with the isotopy g_t from x_i to $g(x_i)$, then finally moving the $g(x_i)$ in a straight line back to the x_i.

Now define

$$\Phi_\mu(g) = \int_{D \times \cdots \times D} \mu(\gamma(g; x_1, \cdots, x_n)) d\text{area}(x_1) \times \cdots \times d\text{area}(x_n)$$

and $\overline{\Phi}_\mu(g) = \lim_{n\to\infty} \frac{1}{n} \Phi_\mu(g^n)$.

LEMMA 5.68. *For any quasimorphism μ on B_n, the function $\overline{\Phi}_\mu$ is a homogeneous quasimorphism on* $\text{Diff}_0^\infty(D, \partial D, \text{area})$.

PROOF. For any two diffeomorphisms g, h, and generic $x_1, \cdots, x_n$ there is equality

$$\gamma(gh; x_1, \cdots, x_n) = \gamma(h; x_1, \cdots, x_n) \cdot \gamma(g; h(x_1), \cdots, h(x_n))$$

in B_n. Homogenizing removes the dependence on the choice of x_i^0. Integrating over $D \times \cdots \times D$ and using the fact that μ is a quasimorphism, we obtain the desired result. □

EXAMPLE 5.69. In case $n = 2$, the group B_2 is isomorphic to $\mathbb{Z}$ and we can take μ to be an isomorphism. In this case, the resulting function $\overline{\Phi}_\mu$ is a *homomorphism* from $\text{Diff}_0^\infty(D, \partial D, \text{area})$ to $\mathbb{R}$, which is equal (after normalization) to the well-known *Calabi homomorphism*.

Calabi [**38**] constructed an invariant for any symplectic diffeomorphism with compact support of a symplectic manifold without boundary. Calabi's construction can be translated into the case of area-preserving diffeomorphisms of the disk as follows. Let θ be a 1-form on D whose exterior derivative $d\theta$ is the area form. If g is an area-preserving diffeomorphism of D fixing ∂D pointwise, then $g^*\theta - \theta$ is closed, and there is a function f on D satisfying $df = g^*\theta - \theta$. The function f is unique up to addition of a constant; normalize f so that it is zero on ∂D. Calabi's homomorphism is defined by the formula

$$\Psi(g) = \int_D f d\theta$$

Changing θ to $\theta' = \theta + dh$ changes f to $f' = f + (h - h^g)$; since g is area-preserving, the integral of $(h - h^g)$ is zero, so Ψ does not depend on the choice of θ.

If g_1 and g_2 are two diffeomorphisms, then

$$(g_1 g_2)^*\theta - \theta = g_2^* g_1^* \theta - g_1^*\theta + g_1^*\theta - \theta$$

so $\Psi(g_1 g_2) = \Psi(g_1) + \Psi(g_2)$. The interpretation of Calabi's homomorphism as an "average braiding number" of pairs of points in the disk is due to Fathi (unpublished); see [**85**].

To define quasimorphisms on $\text{Diff}_0^\infty(S^2, \text{area})$, we need to construct quasimorphisms on $\widehat{B}_n$, the braid group of n-points in the sphere. One way to construct such quasimorphisms is to think of $\widehat{B}_n$ as the mapping class group of a sphere with n punctures, and use the methods of § 3.5, for instance Theorem 3.74. Another, more explicit method is to use the relationship between $\widehat{B}_n$ and B_{n-1}. By thinking of the disk as the once-punctured sphere, one sees that there is a homomorphism $B_{n-1} \to \widehat{B}_n$. The kernel of this map is $\mathbb{Z}$, generated by a "full twist" of all strands; and the image has finite index in $\widehat{B}_n$, and contains the kernel of the permutation map from $\widehat{B}_n$ to the symmetric group S_n. For example, $\widehat{B}_4$ contains the free group F_2 with finite index, and therefore admits an infinite dimensional family of homogeneous quasimorphisms.

Given a (homogeneous) quasimorphism μ on $\widehat{B}_n$, we can construct a homogeneous quasimorphism $\overline{\Phi}_\mu$ on $\text{Diff}_0^\infty(S^2, \text{area})$ as in Definition 5.67. In a similar way Gambaudo–Ghys show ([**86**], Theorem 1.2) that for every closed oriented surface S there exist an infinite dimensional space of homogeneous quasimorphisms on $\text{Diff}_0^\infty(S, \text{area})$.

5.3.3. Higher genus. When S has higher genus, one can construct quasimorphisms on $\text{Diff}_0^\infty(S, \text{area})$ from a hyperbolic structure on S, by a variation of the construction of de Rham quasimorphisms in § 2.3.1.

DEFINITION 5.70 (de Rham quasimorphism). Let S be a closed surface with $\chi(S) < 0$. Fix a hyperbolic structure on S, and let α be a 1-form on S. Given $f \in \text{Diff}_0^\infty(S, \text{area})$, let f_t be an isotopy from id to f. For each $x \in S$, define $\gamma(x, f)$ to be the unique geodesic in S from x to $f(x)$ in the relative homotopy class of the path $f_t(x)$. Then define

$$\phi_\alpha(f) = \int_S \left(\int_{\gamma(x,f)} \alpha \right) \text{darea}$$

LEMMA 5.71. *The function ϕ_α is a quasimorphism on $\text{Diff}_0^\infty(S, \text{area})$ with defect at most $\|d\alpha\|\pi \cdot \text{area}(S)$.*

PROOF. For any point x and any two elements f, g there is a geodesic triangle with edges $\gamma(x,f)$, $\gamma(f(x),g)$, and $\gamma(x,gf)$. By Stokes' theorem,

$$\left| \int_{\gamma(x,f)} \alpha + \int_{\gamma(f(x),g)} \alpha - \int_{\gamma(x,gf)} \alpha \right| \le \|d\alpha\|\pi$$

Now integrate over $x \in S$, and use the fact that f is area-preserving to change variables in the second term on the left hand side. One obtains the estimate

$$|\phi_\alpha(f) + \phi_\alpha(g) - \phi_\alpha(gf)| \le \|d\alpha\|\pi \cdot \text{area}(S)$$

as claimed. □

The homogenizations of ϕ_α are typically nontrivial, and generate an infinite dimensional subspace of Q. When α is a closed 1-form, ϕ_α depends only on the cohomology class $[\alpha] \in H^1(S)$, and is evidently equal to the *flux homomorphism* (Poincaré) dual to $[\alpha]$.

EXAMPLE 5.72 (Ruelle's rotation number [**181**]). The same method does not work directly on $\text{Diff}_0^\infty(T^2, \text{area})$. Nevertheless, Ruelle showed how to define a "rotation quasimorphism" on this group as follows. First, trivialize the tangent bundle; for example, we can choose a Euclidean metric on T^2, and use the flat connection to trivialize the bundle. Given $x \in T^2$ and $f \in \text{Diff}_0^\infty(T^2, \text{area})$, choose an isotopy f_t from id to f. Given a point x, the trivialization lets us canonically identify tangent spaces $T_{f_t(x)}$ and T_x, so we can think of df_t as a path in $\text{GL}(T_x)$. Projectivizing gives a path in $\text{PSL}(T_x)$; lifting to $\widetilde{\text{SL}}(T_x)$ and composing with the rotation quasimorphism defines a number $\rho(x,f)$. A different but homotopic path f_t' determines a homotopic path in $\text{PSL}(T_x)$. Since $\pi_1(\text{Diff}_0^\infty(T^2, \text{area}))$ is generated by loops of translations, $\rho(x,f)$ does not depend on any choices. Now define

$$R(f) = \int_{T^2} \rho(x,f) d\text{area}$$

Similar arguments to those above show that R is a (nontrivial) quasimorphism.

REMARK 5.73. If G is a subgroup of $\text{Diff}_0^\infty(T^2)$ and μ is any G-invariant probability measure on T^2, there is a Ruelle quasimorphism R_μ on G. Similar constructions also make sense on groups of Hamiltonian symplectomorphisms (or on their universal covers) of certain symplectic manifolds.

REMARK 5.74. There is a section from $\text{SL}(2,\mathbb{Z})$ to $\text{Diff}^\infty(T^2, \text{area})$ whose image consists of the linear automorphisms of T^2 fixing a basepoint. This group acts by conjugation on $\text{Diff}_0^\infty(T^2, \text{area})$, and the Ruelle quasimorphism is constant on orbits. Consequently, the Ruelle quasimorphism admits an extension to all of $\text{Diff}^\infty(T^2, \text{area})$.

Also see work of Py, e.g. [**173, 172, 174**] and Entov–Polterovich [**75**] for many more examples of quasimorphisms on various transformation groups.

5.3.4. C^0 case.

The material in this section is taken from [**76**].

The quasimorphisms discussed in § 5.3.2 and § 5.3.3 are evidently continuous in the C^1 topology, and therefore extend continuously to quasimorphisms on groups of the form $\text{Diff}_0^1(S, \text{area})$. If a quasimorphism on $\text{Diff}_0^1(S, \text{area})$ is continuous in the C^0 topology, it extends to $\text{Homeo}_0(S, \text{area})$; this property is more delicate.

The following characterization of continuous quasimorphisms on topological groups is due to Shtern:

THEOREM 5.75 (Shtern [**188**], Thm. 1). *Let G be a topological group. A homogeneous quasimorphism ϕ on G is continuous if and only if it is bounded on some neighborhood of* id.

PROOF. One direction follows from the definition of continuity. Conversely, suppose there is a neighborhood U of id and a constant C so that $|\phi(k)| \le C$ for $k \in U$. For any $g \in G$ and $n \in \mathbb{N}$, define $U(g,n)$ to be the set of $h \in G$ such that $h^n = g^n k$ for some $k \in U$. Evidently, $U(g,n)$ is a neighborhood of g. Moreover, if $h \in U(g,n)$, then by homogeneity,

$$|\phi(h) - \phi(g)| = \frac{1}{n}|\phi(h^n) - \phi(g^n) - \phi(k) + \phi(k)|$$

where $k = g^{-n}h^n$. Since $k \in U$, there is an estimate $|\phi(k)| \le C$. Hence one can estimate

$$|\phi(h) - \phi(g)| \le \frac{1}{n}(D(\phi) + C)$$

Taking n large shows that $\phi(h) \to \phi(g)$ as $h \to g$, so ϕ is continuous. □

Using this characterization, Entov–Polterovich–Py derive the following theorem in the context of transformation groups. Given a surface S, let $\text{Ham}(S, \text{area})$ denote the subgroup of $\text{Diff}_0^\infty(S, \text{area})$ consisting of Hamiltonian diffeomorphisms (i.e. those in the kernel of every flux homomorphism).

THEOREM 5.76 (Entov–Polterovich–Py). *Let ϕ be a homogeneous quasimorphism on* $\text{Ham}(S, \text{area})$. *Then ϕ is continuous in the C^0 topology if and only if there is some positive constant a so that if $D \to S$ is any embedded disk of area at most a, then ϕ vanishes identically on the subgroup $G(D)$ of elements supported in D.*

PROOF. We give the sketch of a proof; for details, see [**76**]. Suppose ϕ is continuous, and let U be a neighborhood of id (in the C^0 topology) for which there is a constant C as in the conclusion of Theorem 5.75. If D_0 is sufficiently small in diameter, then $G(D_0) \subset U$, and therefore ϕ is bounded on $G(D_0)$. But since ϕ is a homogeneous quasimorphism, and $G(D_0)$ is a group, ϕ must vanish identically on $G(D_0)$. Now, if D is any other disk with $\text{area}(D) \le \text{area}(D_0)$, there is an area-preserving Hamiltonian isotopy from D to D_0. Hence $G(D_0)$ and $G(D)$ are conjugate, and the conclusion follows.

Conversely, suppose there is a positive constant a with the desired properties. There is a neighborhood U of the identity so that S can be covered with finitely many disks D_i for $i \le N$, each of area at most a, so that any $f \in U$ can be written as a product $f = g_1 g_2 \cdots g_N$ where the support of each g_i is contained in D_i (and therefore $g_i \in G(D_i)$). Since ϕ vanishes identically on each $G(D_i)$, the value of ϕ on f is bounded by $(N-1)D(\phi)$, and therefore ϕ is continuous, by Theorem 5.75. □

A homogeneous quasimorphism on $\text{Diff}_0^\infty(S, \text{area})$, continuous on $\text{Ham}(S, \text{area})$, and linear on every one-parameter subgroup, is continuous in the C^0 topology, and therefore extends to $\text{Homeo}_0(S, \text{area})$.

REMARK 5.77. The most delicate aspect of Theorem 5.76 is the fragmentation lemma (i.e. to show that one can express a Hamiltonian diffeomorphism sufficiently C^0 close to the identity as a product of boundedly many diffeomorphisms supported in small disks). This depends on work of Le Roux [**133**]. Note that the assumption that the diffeomorphism be Hamiltonian is essential.

EXAMPLE 5.78. When the genus of S is large, the homogenizations of the de Rham quasimorphisms (Definition 5.70) vanish on $G(D)$ for *any* embedded disk D. Hence they are continuous in the C^0 topology, and extend to quasimorphisms on $\mathrm{Homeo}_0(S, \mathrm{area})$.

It is still unknown whether $\mathrm{Homeo}_0(S^2, \mathrm{area})$ admits any nontrivial quasimorphism.

REMARK 5.79. The study of quasimorphisms on (mostly 2-dimensional) transformation groups is an active and fertile area. In addition to the work of Entov–Polterovich [**75**] and Gambaudo–Ghys referred to above, we mention only the survey [**169**] by Polterovich, and [**175**] by Py, discussing relations of this material to Zimmer's program.

CHAPTER 6

Combable functions and ergodic theory

In this chapter we study quasimorphisms on hyperbolic groups, especially counting quasimorphisms, from a *computational* perspective. We introduce the class of *combable functions* (and the related classes of weakly combable and bicombable functions) on a hyperbolic group, and show that the Epstein–Fujiwara counting functions are bicombable.

Conversely we show that bicombable functions satisfying certain natural conditions are quasimorphisms; thus quasimorphisms and bounded cohomology arise naturally in the study of automatic structures on hyperbolic groups, a fact which might at first glance seem surprising.

The (asymptotic) distribution of values of a combable function may be described very simply using stationary Markov chains. Consequently, we are able to derive a central limit theorem for the distribution of values of counting quasimorphisms on hyperbolic groups.

The main reference for this section is Calegari–Fujiwara [**50**], although Picaud [**166**] and Horsham–Sharp [**113**] are also relevant.

6.1. An example

6.1.1. Random walk on $\mathbb{Z}$.

Definition 6.1. A sequence of integers $x = (x_0, x_1, \cdots)$ is a *walk* on $\mathbb{Z}$ if it satisfies the following two properties:

(1) (initialization) $x_0 = 0$
(2) (unit step) for all $n > 0$, there is an equality $|x_n - x_{n-1}| = 1$

The *length* of a walk x is *one less* than the number of terms in the sequence x. So, for example, $(0, 1, 2)$ has length 2, while $(0, 1, 0, -1, -2)$ has length 4.

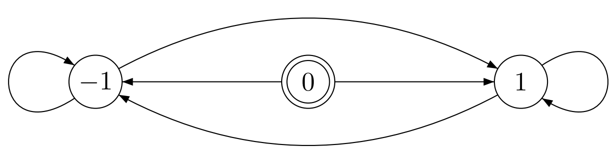


Figure 6.1. Walks on $\mathbb{Z}$ of length n are in bijection with walks on Γ of length n.

Knowing the successive differences $x_n - x_{n-1} \in \{-1, 1\}$ determines x, so there is a bijection between walks of length n, and strings of length n in the alphabet

$\{-1, 1\}$. This correspondence may be encoded graphically as follows. Let Γ be the directed graph depicted in Figure 6.1. A walk x on $\mathbb{Z}$ "determines" a corresponding walk x' on Γ starting at the initial vertex (labeled 0) where the labels on the vertices in the itinerary of x' are exactly the sequence of successive differences $x_n - x_{n-1}$. In other words,

$$x'_n = x_n - x_{n-1}$$

Formally, x' is a kind of *discrete derivative* of x. The advantage of the correspondence $x \to x'$ is that it replaces a random walk on an infinite (but homogeneous) graph (i.e. $\mathbb{Z}$) with a random walk on a *finite* graph.

Let X_n denote the set of walks on $\mathbb{Z}$ of length n, and let $v : X_n \to \mathbb{Z}$ be the function which takes each walk to the last integer in the sequence. For example, $v(0, 1, 2, 1) = 1$ and $v(0, -1, -2, -3, -2) = -2$.

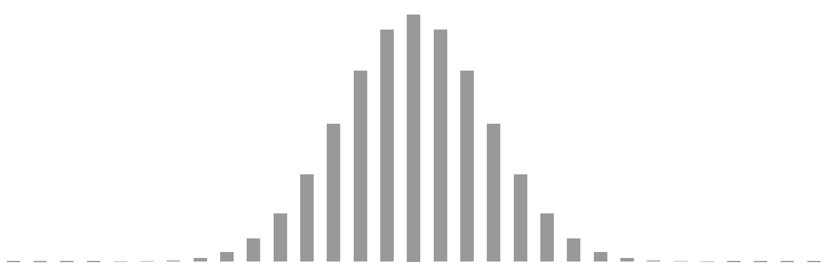


FIGURE 6.2. histogram showing the frequency of outcomes for all walks of length 30 on $\mathbb{Z}$

There are 2^n walks of length n. The set of values of v on X_n are the integers of the form $2i - n$ for $0 \le i \le n$, and the number of elements of X_n taking the value $2i - n$ is $\binom{n}{i} = \frac{n!}{(n-i)!i!}$. A histogram of this data for the case $n = 30$ is contained in Figure 6.2.

This figure has some significant qualitative features: left-right symmetry, the fact that all realized values have the same parity, and so forth. Most notable are the long flat tails on either side. If we rescale the graph horizontally by a factor of n^{-1}, and vertically so that the total area under the graph is equal to 1, the distribution becomes more and more peaked and "limits" to a Dirac distribution with all the mass centered at the origin (technically, this is convergence in the sense of distribution). However, if we instead rescale the graph horizontally by a factor of $n^{-1/2}$, the distribution converges to the familiar "bell curve", or Gaussian. If we let $\bar{v}_n$ denote the value of v on a random element of X_n (with the uniform distribution), then $\bar{v}$ is not a number but rather a (discrete) probability measure on $\mathbb{R}$. The Central Limit Theorem for binomial distributions (see [**96**], Thm. 9.1) says that there is convergence in the sense of distribution

$$\lim_{n\to\infty} \mathbf{P}(s \le n^{-1/2}\bar{v}_n \le t) = \frac{1}{\sqrt{2\pi}} \int_s^t e^{-x^2/2} dx$$

where $\mathbf{P}(\cdot)$ denotes probability, and $s \le t$ are any two real numbers.

6.1.2. Random value of a homomorphism. Given a group G and a function $f : G \to \mathbb{R}$ it is natural to ask how the values of f are distributed on G. If G is finitely generated, we can study *statistical properties* of the values of f on the set of elements of G of (word) length n, as a function of n. This analysis will be most informative when the function f is adapted to the geometry and algebra of G; the most important case therefore is when f is a homomorphism.

In order to keep the discussion concrete, we restrict attention in what follows to free groups. Let F denote the free group generated by two elements a, b, and let $\rho : F \to \mathbb{Z}$ be the unique homomorphism which sends a to 1 and b to 0 (writing $\mathbb{Z}$ additively). A basic question is to ask what is the distribution of the values of ρ on the group F.

If we take $S = \{a, b, a^{-1}, b^{-1}\}$ to be a symmetric generating set for F, the Cayley graph $C_S(F)$ is an infinite regular 4-valent tree. Let γ denote a geodesic in $C_S(F)$ starting at id, and let $\rho(\gamma)$ denote the corresponding walk in $\mathbb{Z}$, whose itinerary consists of the values of ρ on successive vertices of γ. As in the case of a random walk on $\mathbb{Z}$, the situation is clarified by considering, in place of $\rho(\gamma)$, the *discrete derivative*; i.e. by considering how the value of ρ *changes* on successive vertices of γ.

6.1.3. Digraphs. Every element of F is represented by a unique reduced word in the generators, corresponding to the unique geodesic in $C_S(G)$ starting at id and with a given endpoint. Reduced words are certified by *local data*: a word is reduced if and only if no a follows or precedes an a^{-1}, and if no b follows or precedes a b^{-1}. Let S^* denote the set of all finite words in the generating set S, and let W_n denote the set of reduced words in S^* of length n. Let $W = \cup_n W_n$.

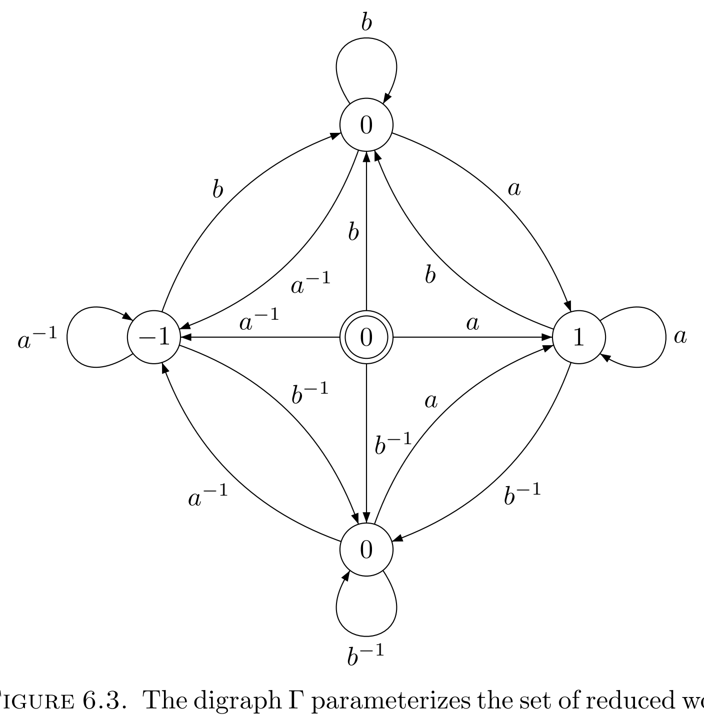


FIGURE 6.3. The digraph Γ parameterizes the set of reduced words in F

Elements of F are in bijection with elements of W by taking each element to the unique reduced word which represents it. Moreover, elements of W are in bijection with certain walks on a *directed graph* Γ, depicted in Figure 6.3 (ignore the numbers on the vertices for the moment). There is a special *initial vertex* with no incoming edges, and four other vertices which have both incoming and outgoing edges. In computer science and combinatorics, a directed graph is usually called a *digraph*, and we use this terminology in what follows. If we need to stress that a particular digraph has an initial vertex, we call it a *pointed digraph*. So Γ in Figure 6.3 is a (pointed) digraph.

A *reduced* word $w \in W$ determines a directed path in Γ starting at the initial vertex, by reading the letters one by one (from left to right) and traversing at each stage the edge of Γ labeled by the corresponding letter of w. Conversely, a directed path in Γ starting at the initial vertex determines a reduced word, determined by the string consisting of the edge labels visited in the path. Under this bijection, elements of W_n correspond to directed paths in Γ of length n.

The information in a digraph can be encoded in the so-called *adjacency matrix*.

Definition 6.2. Let Γ be a digraph with vertices v_i. The *adjacency matrix* of Γ is the square matrix whose entries are determined by the formula

$$M_{ij} = \begin{cases} 1 \text{ if there is a directed edge from } v_i \text{ to } v_j \\ 0 \text{ otherwise} \end{cases}$$

Spectral properties of M reflect geometric properties of Γ. The most explicit example of this is the following Lemma, which says that directed paths in Γ are counted by the entries of powers of M.

Lemma 6.3. *For any n and any vertices v_i, v_j the number of directed paths in Γ from v_i to v_j of length n is $(M^n)_{ij}$.*

Proof. We prove the statement by induction. It is tautologically true for paths of length 1, so assume it is true for paths of length $n-1$. By induction, for any v_k there are $(M^{n-1})_{ik}M_{kj}$ paths of length n from v_i to v_j whose penultimate vertex is v_k. Summing over k gives the desired result. □

The following topological property of digraphs is the analogue of irreducibility in the algebraic context.

Definition 6.4. A digraph is *recurrent* if there is a directed path from any vertex to any other vertex.

By Lemma 6.3, a digraph is recurrent if and only if for any i, j there is some n (which may depend on i, j) for which $(M^n)_{ij}$ is positive.

Definition 6.5. A matrix with non-negative entries is a *Perron–Frobenius matrix* if for any i, j there is some n for which $(M^n)_{ij}$ is positive.

The graph Γ of Figure 6.3 is not recurrent, but the subgraph Γ' consisting of vertices and edges disjoint from the initial vertex is recurrent. Let M be the adjacency matrix of Γ', so

$$M = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

Since Γ' is recurrent, the matrix M is a Perron–Frobenius matrix. In fact, every entry of M^n is positive whenever $n \geq 2$.

For such matrices, one has the fundamental Perron–Frobenius Theorem:

THEOREM 6.6 (Perron–Frobenius). *Let M be a real non-negative matrix so that every entry of M^n is positive for some $n > 0$. Then the following statements hold.*

(1) *M has a positive real eigenvalue λ. Every other eigenvalue ξ satisfies $|\xi| < \lambda$.*
(2) *The algebraic and geometric multiplicities of λ are both equal to 1.*
(3) *There are left and right eigenvectors of M with eigenvalue λ, spanning their respective 1-dimensional eigenspaces, with positive entries.*

See for example [**11**] for a proof. A matrix M with the property above is sometimes called *regular*.

If M is symmetric, the left and right λ-eigenvectors of M are transposes of each other, but for general M this need not be the case.

For the case of Γ' as above, the matrix M is symmetric, and the vector $v = (1/4, 1/4, 1/4, 1/4)$ (resp. v^T) is a left (resp. right) eigenvector for M with eigenvalue 3 and L^1 norm equal to 1.

If M is merely non-negative (with no assumption that there is a power all of whose entries are strictly positive), the situation is more complicated. Since we will need to study this case in the sequel, we state the following proposition.

PROPOSITION 6.7 (Weak Perron–Frobenius). *Let M be a real non-negative matrix. Then M has a positive real eigenvalue λ with left and right eigenvectors, and every other eigenvalue ξ satisfies $|\xi| \le \lambda$.*

If for every i, j there is an n (possibly depending on i, j) for which $(M^n)_{ij}$ is positive, then every eigenvalue ξ with $|\xi| = \lambda$ has the form $\omega\lambda$ for some root of unity ω. Moreover, for every ξ with $|\xi| = \lambda$ the algebraic and geometric multiplicities of ξ are equal, and there are left and right λ eigenvectors for ξ with positive entries.

See [**11**]. A matrix with the property that for all i, j there is n depending on i, j such that $(M^n)_{ij}$ is positive, is sometimes said to be *ergodic* or *irreducible*. An ergodic matrix which is not regular is sometimes called *cyclic*.

To say that the algebraic and geometric multiplicities of an eigenvalue ξ are equal just means that the Jordan block of the eigenvalue ξ is diagonal; i.e. that the generalized ξ-eigenspace is a genuine eigenspace. The weak Perron–Frobenius Theorem can be deduced from the (ordinary) Perron–Frobenius Theorem by approximating a non-negative matrix by a positive matrix.

6.1.4. Random walks on Γ'. For each integer $n \ge 0$, let X_n denote the set of walks on Γ' of length n starting at any vertex. For each $m < n$ there is a *prefix* function $p_{n,m} : X_n \to X_m$ which just forgets the last $n - m$ terms in the sequence. Each map $X_n \to X_{n-1}$ is finite to one. The inverse limit

$$X := \lim_{\leftarrow} X_n$$

is topologically a Cantor set, and parameterizes the set of right-infinite walks on Γ'. We write a typical $x \in X_n$ as a finite sequence $x = (x_0, x_1, \cdots, x_n)$ and an element $x \in X$ as an infinite sequence $x = (x_0, x_1, \cdots)$. It comes together with prefix maps $p_n : X \to X_n$ satisfying $p_{n,m} p_n = p_m$ for all $m < n$.

DEFINITION 6.8. The *shift* map $S : X \to X$ takes a walk to the suffix consisting of all but the first vertex. In co-ordinates,

$$S(x_0, x_1, \cdots) = (x_1, x_2, \cdots)$$

Definition 6.9. A *cylinder* is an open subset of X determined by fixing a finite number of the co-ordinates x_i of an element x.

Let $\mathcal{B}$ denote the σ-algebra on X generated by all cylinders. Note that $\mathcal{B}$ is the Borel σ-algebra on X associated to its natural inverse limit topology.

The shift map S acts continuously on X, and therefore measurably with respect to $\mathcal{B}$. Any measurable map on a compact space preserves some probability measure. In our example, there is a unique probability measure μ on X which is invariant under S, with the property that for all n, the pushforward $(p_n)_*\mu$ is equal to the uniform probability measure on X_n.

If $\pi : X \to \Gamma'$ takes an element to its initial vertex, and $x \in X$ is chosen at random, the sequence

$$\pi(x), \pi(Sx), \pi(S^2x), \cdots$$

is an infinite random walk on Γ', where the transition probabilities to move from vertex to vertex at each stage are given by the matrix

$$N = \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{pmatrix}$$

N is a *stochastic matrix*, meaning that the entries are non-negative, and the vector $\mathbf{1} := (1, \cdots, 1)^T$ is a right eigenvector with eigenvalue 1. The uniform probability measure on Γ' is stationary for N, meaning that $\mathbf{1}^T$ is a left eigenvector with eigenvalue 1. Again, in general, a left eigenvector for a stochastic matrix will *not* correspond to the uniform measure, however a stationary measure exists by the Perron–Frobenius Theorem 6.6.

The essential property of the process $(x_0, x_1, \cdots)$ corresponding to a random $x \in X$ (with respect to the uniform measure) is that for each i, the probability that x_{i+1} will be in a given state depends only on x_i, and not on x_j for any $j < i$. Informally, this can be summarized by saying that *future states depend only on the present, and are independent of the past.* This property of a random process is generally called the *Markov property*, and the usual terminology for this is that a random walk on Γ' is (governed by) a *stationary Markov chain.* The Perron–Frobenius property of the transition matrix N is summarized by saying that this Markov chain is *ergodic*.

For each n, let Y_n be the subspace of X_n consisting of walks that begin at the initial vertex. Elements of Y_n are in bijection with elements of F of word length n. Each element of Y_n corresponds to a cylinder in X consisting of infinite walks that begin with a given prefix. The measure μ induces in this way a measure on each Y_n; after scaling, this is the uniform measure in which each element has probability $1/(4 \cdot 3^{n-1})$. The homomorphism ρ determines a function $d\rho$ from Γ' to $\mathbb{Z}$ by the formula

$$d\rho(\mathbf{s}(ws)) = \rho(ws) - \rho(w)$$

where $\mathbf{s} : Y_n \to \Gamma'$ sends a (finite) walk to its terminal vertex. In other words, the function $d\rho$ measures how much the value of ρ *changes* on the increasing prefixes of a reduced word. If a vertex v_i of Γ' is encoded as a column vector, the function $d\rho$ can be encoded as a row vector of the same length, and evaluation of the function amounts to contraction of vectors. In our example, $d\rho$ is the vector $(1, 0, -1, 0)$.

Let $\bar{S}_n$ be a random variable whose value is

$$\bar{S}_n = \sum_{i=1}^{n} d\rho(x_i)$$

where $x = (x_0, x_1, \cdots, x_n)$ is a random element of Y_n. In other words, $\bar{S}_n$ is the value of ρ on a random element of F of word length n.

Technically, $\bar{S}_n$ should be thought of as a probability measure on $\mathbb{R}$, supported in $\mathbb{Z}$. Every $x \in Y_n$ determines an integer $\sum d\rho(x_i)$, and this determines a map $Y_n \to \mathbb{Z}$. The (uniform) measure on Y_n pushes forward under this map to a measure on $\mathbb{Z}$, which by definition is $\bar{S}_n$.

The Central Limit Theorem for ergodic stationary Markov chains (see [**179**] p. 231) says that there is a convergence in the sense of distribution

$$\lim_{n\to\infty} \mathbf{P}\left(s \le \frac{\bar{S}_n - nE}{\sqrt{\sigma^2 n}} \le t\right) = \frac{1}{\sqrt{2\pi}} \int_s^t e^{-x^2/2} dx$$

where E is the mean of $d\rho$ on Γ' with respect to the stationary measure (which is equal to 0 in this case) and σ^2 is an algebraic number which can be determined from N, μ and $d\rho$.

6.1.5. More complicated examples. The homomorphism ρ in the example above is a very simple example of a big counting quasimorphism; explicitly, $\rho = H_a$ in the notation of Definition 2.25. We would like to study the distribution of H_w on F for an arbitrary reduced word $w \in F$. The problem is that the digraph Γ defined in the last section is not adequate for our purpose. A reduced word in F determines a walk in Γ, but the vertex at each step only "remembers" one letter at a time. In order to count occurrences of a word w or its inverse w^{-1} we need a more complicated digraph whose vertices remember enough information to keep track of each occurrence of w or w^{-1}.

Definition 6.10. Let Γ be a pointed digraph. Define $\Gamma_0 = \Gamma$. For each $n > 0$, define inductively a pointed digraph Γ_n as follows.

The vertices of Γ_n consist of an initial vertex, together with one vertex for every directed path in Γ_{n-1} of length 1 (with any starting vertex). The edges of Γ_n (except for those which start at the initial vertex) correspond to pairs of composable paths; i.e. pairs of paths of length 1 which can be concatenated to form a path of length 2.

Finally, for every path of length 1 in Γ_{n-1} starting at the initial vertex, add a directed edge in Γ_n from the initial vertex to the corresponding vertex of Γ_n.

Γ_n is called the nth *refinement* of Γ.

Remark 6.11. The construction of a refinement makes sense for any pointed digraph.

Remark 6.12. Notice that each Γ_n is finite if Γ is, and contains a unique maximal recurrent subgraph Γ'_n if Γ does.

See Figure 6.4 for an example of the first refinement Γ_1, where Γ is the example from Figure 6.3. For the sake of legibility, labels on the arrows (which are elements of the generating set S) have been suppressed. Note how complicated this example is, with 17 vertices and 52 edges. In general, the graph Γ_n contains $O(\lambda^n)$ vertices and $O(\lambda^{n+1})$ edges, where λ is the Perron–Frobenius eigenvalue of the transition matrix of Γ, so actually constructing Γ_n is typically not practical, even for moderate values of n.

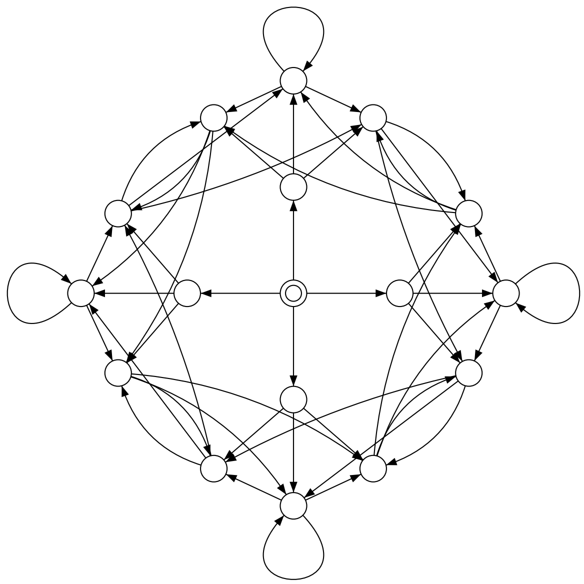


FIGURE 6.4. The digraph Γ_1 is the (first) refinement of Γ. dH_{ab} is a function from the states of Γ_1 to $\mathbb{Z}$.

By induction, the stationary measure for each Γ_n is the uniform measure on the subgraph Γ'_n, and the transition matrix has equal probability for each edge of Γ'_n.

For any $n, m \geq 0$ there is an equality $(\Gamma_n)_m = \Gamma_{n+m}$. Moreover, by induction, $\Gamma'_n = (\Gamma')_n$. Vertices in Γ'_n correspond to paths of length n in Γ'. Let $w \in F$ be a reduced word of length n, and let $H_w = C_w - C_{w^{-1}}$ be the big counting quasimorphism. Define $dH_w : \Gamma'_n \to \mathbb{Z}$ by setting dH_w equal to 1 on the vertex corresponding to the path w in Γ', and -1 on the vertex corresponding to the path w^{-1} in Γ'. The Central Limit Theorem for ergodic stationary Markov chains implies the following theorem.

THEOREM 6.13 (Calegari–Fujiwara). *Let H_w be a big counting quasimorphism on a free group. If $\bar{H}_w(n)$ denotes the value of H_w on a random word in F of length n (in a standard symmetric generating set), then there is convergence in the sense of distributions*

$$n^{-1/2}\bar{H}_w(n) \to N(0, \sigma)$$

for some σ depending on w.

It is one of the goals of this chapter to generalize this theorem to a broader class of quasimorphisms on arbitrary word hyperbolic groups.

6.1.6. Hölder quasimorphisms. The property of big counting quasimorphisms described in Theorem 6.13 holds for other interesting classes of quasimorphisms on free groups, including those with the so-called *Hölder property*.

DEFINITION 6.14. For any $g \in F$, and any function ψ on F, define

$$\Delta_a\psi(g) = \psi(g) - \psi(ag)$$

For $x, y \in F$ let $(x|y)$ denote the Gromov product; i.e.

$$(x|y) = (|x| + |y| - |x^{-1}y|)/2$$

In other words, $(x|y)$ is the length of the biggest common prefix of the words x, y.

Say that a quasimorphism $\psi \in Q(F)$ is *Hölder* if for any $a \in F$ there are constants $C, c > 0$ such that for any $x, y \in F$ there is an inequality

$$|\Delta_a\psi(x) - \Delta_a\psi(y)| \le Ce^{-c(x|y)}$$

Note that the constants C, c depend on a, but not on x or y.

Horsham and Sharp [**113**], extending some results in Matthew Horsham's PhD thesis, prove the following theorem:

THEOREM 6.15 (Horsham–Sharp). *Let ψ be a Hölder quasimorphism on a free group. If $\bar{\psi}(n)$ denotes the value of ψ on a random word in F of length n (in a standard symmetric generating set), then there is convergence in the sense of distributions*

$$n^{-1/2}\bar{\psi}(n) \to N(0, \sigma)$$

for some σ.

The argument involves (nonstationary) Markov chains obtained from subshifts of finite type, and the associated thermodynamic formalism. These results can also be generalized to surface groups.

Big counting quasimorphisms are trivially seen to be Hölder, since $\Delta_a\psi(x) = \Delta_a\psi(y)$ whenever $(x|y)$ is bigger than $|a|$. But small counting quasimorphisms are not, as the following example (from [**50**]) shows.

EXAMPLE 6.16. Let $h := h_{abab}$. Then

$$h(\underbrace{babab\cdots ab}_{4n+1}) = n, \quad h(\underbrace{ababab\cdots ab}_{4n+2}) = n$$

but

$$h(\underbrace{babab\cdots ab}_{4n+3}) = n, \quad h(\underbrace{ababab\cdots ab}_{4n+4}) = n+1$$

Although small counting quasimorphisms are not Hölder, they nevertheless have a great deal in common with big counting quasimorphisms: both are examples of *bicombable functions*, to be defined in § 6.3.2. Ultimately, we will prove a version of the Central Limit Theorem valid for all bicombable functions on arbitrary word-hyperbolic groups.

6.1.7. Rademacher function. There are natural ways to filter elements in free groups other than by word length. If one thinks of a (virtually) free group as the fundamental group of a cusped hyperbolic surface (orbifold), it is natural to count conjugacy classes (which correspond to closed geodesics) and sort them by geodesic length. The noncompactness of the surface leads to quite distinctive features of the theory. In this context, we mention a result of Peter Sarnak, showing that the Rademacher function on conjugacy classes in the group $\mathrm{PSL}(2, \mathbb{Z})$ has values which obey a Cauchy distribution, in contrast to the Gaussian distributions discussed above.

Ghys [**91**] gave an elegant topological definition of the Rademacher function. The group $\mathrm{PSL}(2, \mathbb{Z})$ acts on the hyperbolic plane $\mathbb{H}^2$ by isometries, with quotient the $(2, 3, \infty)$-triangle orbifold Δ. Each element A of $\mathrm{PSL}(2, \mathbb{Z})$ whose trace has absolute value > 2 fixes a unique axis in $\mathbb{H}^2$, which covers a geodesic in Δ. This geodesic lifts to an embedded loop γ_A in the unit tangent bundle $UT\Delta$ which is homeomorphic to the quotient $\mathrm{PSL}(2,\mathbb{R})/\mathrm{PSL}(2,\mathbb{Z})$. As is well known, $UT\Delta$ is homeomorphic to the complement of the trefoil knot T in S^3.

DEFINITION 6.17. For $A \in \mathrm{PSL}(2,\mathbb{Z})$ with $|\mathrm{tr}(A)| > 2$, define $R(A)$ to be the linking number of γ_A and T in S^3.

Ghys relates $R(A)$ to the classical Rademacher function, which is defined in terms of Gauss sums, and is intimately related to the Dedekind η function. If we think of $\mathrm{PSL}(2,\mathbb{Z})$ as a subgroup of $\mathrm{Homeo}^+(S^1)$, and $\widetilde{\mathrm{SL}}(2,\mathbb{Z})$ as its preimage in $\mathrm{Homeo}^+(\mathbb{R})^{\mathbb{Z}}$, then there is a rotation quasimorphism rot on $\widetilde{\mathrm{SL}}(2,\mathbb{Z})$. Let $\rho : \widetilde{\mathrm{SL}}(2,\mathbb{Z}) \to \mathbb{Z}$ be the unique homomorphism that takes the value 6 on the generator of the center (i.e. the element that acts on $\mathbb{R}$ as $z \to z+1$). Then $6 \cdot \mathrm{rot} - \rho$ descends to the quasimorphism R on $\mathrm{PSL}(2,\mathbb{Z})$.

Conjugacy classes of elements A in $\mathrm{PSL}(2,\mathbb{Z})$ with $|\mathrm{tr}| > 2$ correspond to closed geodesics γ_A in Δ. Let $|\gamma_A|$ denote the length of γ_A. For each real number y, define

$$\pi(y) := \#\{A : |\gamma_A| \le y\}$$

The behavior of $\pi(y)$ for large y is known; in fact,

$$\pi(y) = \mathrm{Li}(e^y) + O(e^{7y/10})$$

where

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x}$$

Sarnak shows that the Rademacher function R, filtered by geodesic length, satisfies a Cauchy distribution; i.e.

THEOREM 6.18 (Sarnak). *With notation as above,*

$$\lim_{y\to\infty} \frac{1}{\pi(y)} \#\{A : |\gamma_A| \le y \text{ and } a \le \frac{R(A)}{|\gamma_A|} \le b\} = \frac{1}{\pi}\left(\arctan(\frac{b\pi}{3}) - \arctan(\frac{a\pi}{3})\right)$$

The "reason" for the difference in observed distributions has to do with the relationship between word length and geodesic length in $\mathrm{PSL}(2,\mathbb{Z})$. The group $\mathrm{PSL}(2,\mathbb{Z})$ is virtually free, containing a subgroup Γ of index 12 which is isomorphic to F_2. The surface Δ is non-compact, with a cusp. A geodesic γ_A which winds a lot around the cusp might have length as small as $O(\log(n))$ where n is the word length of A. If w is a reduced word in F_2 of the form $a^{n_1}b^{n_2}\cdots a^{n_k}b^{n_k}$ then the length of w is $\sum n_i$ but the length of the geodesic γ_w is $O(\sum \log(n_i))$. Since quasimorphisms are homomorphisms on cyclic subgroups, such a word probably has an unusually large value of R for its word length, and especially for its geodesic length, thus giving rise to the fat tails of the Cauchy distribution.

6.2. Groups and automata

Our analysis in § 6.1 depended crucially on the fact that elements in a free group could be parameterized by directed paths in a digraph, namely the digraph Γ from Figure 6.3 and its refinements. The proper generalization of this fact for more complicated groups involves the theory of combings and regular languages.

6.2.1. Regular languages. Let S be a finite alphabet, and let S^* denote the set of all (finite) words in the alphabet S.

DEFINITION 6.19. A *language* is a subset $L \subset S^*$. A language is *prefix closed* if every prefix of an element of the language is also in the language.

DEFINITION 6.20. A *finite state automaton* on a fixed alphabet is a digraph with a distinguished initial vertex (the *input state*), and with oriented edges labeled by letters of the alphabet, such that at each vertex there is at most one outgoing edge with any label.

The vertices are also called the *states* of an automaton. A word $w \in S^*$ determines a directed path in the automaton, which starts at the initial vertex at time 0, and moves along a directed edge labeled w_i at time i, if one exists, or halts if not. The resulting path in the automaton is said to be obtained by *reading* the word w.

Some subset of vertices are labeled *accept* states. If the automaton reads to the end of w without halting, the last vertex of the path is the *final state* and the word is *accepted* if the final state is an accept state, and rejected otherwise.

DEFINITION 6.21. A *regular language* is the set of words in some fixed alphabet accepted by some finite state automaton.

REMARK 6.22. For regular languages which are prefix closed, one can restrict attention to automata in which every state is an accept state. In the sequel we shall be exclusively interested in prefix closed regular languages, and therefore every state in our automata will be an accept state.

The concept of a finite state automaton or a regular language is best understood by considering some simple examples.

EXAMPLE 6.23. Let $S = \{a, b\}$. The following languages are regular:

(1) The set of all words in S^*
(2) The set of all words in S^* which contain the string *baa* but not the string *abba*
(3) The set of all words in S^* with at least 5 a's
(4) The set of all words in S^* for which the number of a's and b's have different parities

The following languages are not regular:

(1) The set of all words of the form $a^n b^n$
(2) The set of all palindromic words
(3) The set of all words with prime length
(4) The set of all words which contain more a's than b's

In words, a finite state automaton is a machine with a finite amount of memory. It reads the letters of w in order, and cannot go back and re-read some subword. In practice, automata can be described informally in terms of the task they perform, rather than explicitly in terms of vertices and edges.

Suppose L is regular and prefix closed. Then there is a finite state automaton A which accepts L and for which every vertex is an accept state. The underlying digraph Γ of the automaton A *parameterizes* L, in the sense that there is a natural bijection

$$\text{directed paths in } \Gamma \text{ starting at the initial vertex} \longleftrightarrow \text{elements of } L$$

NOTATION 6.24. Suppose Γ parameterizes L. Let $w \in L$ and as above, let w_i denote the ith letter of w. We let $\gamma_i(w)$ denote the ith vertex of the corresponding path in Γ, respectively γ_i if w is understood, and let $\gamma(w)$ (resp. γ) denote the endpoint of the path in Γ.

WARNING 6.25. For a fixed regular, prefix closed language L there are *many* digraphs Γ which parameterize L. For instance, if Γ parameterizes L, then so does every resolution Γ_n.

6.2.2. Combings.

NOTATION 6.26. If G is a group and S is a generating set, there is a natural evaluation map $\mathsf{e} : S^* \to G$ taking a word in the generators to the element in G it represents. Sometimes, where no confusion can arise, we omit e, so that the same symbol w may represent a word in S^* or an element of G. If w is a word in S^*, we let $\mathsf{e}_i(w)$ denote the path in G whose ith element is the image under e of the prefix of w of length i.

DEFINITION 6.27. Let G be a group with finite symmetric generating set S. A *combing* of G with respect to S is a regular language $L \subset S^*$ which satisfies the following conditions:

(1) The evaluation map $\mathsf{e} : L \to G$ is a bijection
(2) L is prefix closed
(3) L is geodesic; i.e. elements of L represent geodesic paths in $C_S(G)$

WARNING 6.28. Definitions of combings differ in the literature. All three bullets in Definition 6.27 (and sometimes even the condition that L is regular) are omitted or modified by some authors!

Let L define a combing of G with respect to S, and let Γ be a digraph which parameterizes L. Every path in Γ determines a path in $C_S(G)$ starting at the identity. The conditions in Definition 6.27 imply that the union of these paths is an isometrically embedded maximal spanning tree in $C_S(G)$.

One of the principal motivations for studying combings is the following theorem, first proved by Cannon (though he used different terminology):

THEOREM 6.29 (Cannon [**51**], [**77**]). *Let G be a word-hyperbolic group, and S a finite symmetric generating set. There is a combing of G with respect to S.*

In fact, many natural, explicit combings exist. Choose a total ordering $\prec$ on the elements of S. This induces a *lexicographic* ordering (i.e. a dictionary ordering) on the elements of S^*. The language L of lexicographically first geodesic words in S^* satisfies the bullet conditions of Definition 6.27; the main content of Theorem 6.29 is that L is regular.

6.3. Combable functions

6.3.1. Left and right invariant Cayley metrics. Let G be a group with finite symmetric generating set S. There are *two* natural metrics on G associated to S — a *left* invariant metric d_L which is just the metric induced by the usual path metric in the Cayley graph $C_S(G)$, and a *right* invariant metric, where $d_R(a,b) = d_L(a^{-1}, b^{-1})$. If $|\cdot|$ denotes the word length of an element in G, then

$$d_L(a,b) = |a^{-1}b|, \quad d_R(a,b) = |ab^{-1}|$$

Each metric d_L, d_R is induced from a path metric. The geometry of a metric space X, d_X may be probed effectively by studying the space of all Lipschitz functions $X \to \mathbb{R}$. For G a group, it is natural to probe G by functions which are Lipschitz with respect to either the d_L or d_R metric, or both simultaneously.

Note that a function $f : G \to \mathbb{Z}$ is Lipschitz for the d_L metric if and only if there is a constant C so that for all $a \in G$ and all $s \in S$,

$$|f(as) - f(a)| \le C$$

Similarly, f is Lipschitz for the d_R metric if

$$|f(sa) - f(a)| \le C$$

The properties of being Lipschitz for d_L or d_R respectively do not depend on a choice of generating set for S (but the constants will).

Remark 6.30. It is psychologically challenging to find a good way to perceive a group G simultaneously in both its d_L and d_R metrics. An analogy is the relationship between matrices and rooted trees. The elements of a matrix can be thought of as the leaves of a depth 2 rooted tree in two distinct ways. The depth 1 nodes can either be thought of as denoting *rows* or as *columns*. The two tree structures are obtained by thinking of the index sets as affine spaces for the action of a group $\mathbb{Z}$, and the two different tree structures correspond to the actions of $\mathbb{Z}$ from the left and from the right.

Any homomorphism $G \to \mathbb{Z}$ is Lipschitz in both the d_L and d_R metrics. But hyperbolic groups do not always admit many (or even any) homomorphisms to $\mathbb{Z}$ (for instance, fundamental groups of quaternionic hyperbolic manifolds have Kazhdan's property (T), and therefore no subgroup of finite index admits a homomorphism to $\mathbb{Z}$). However, quasimorphisms are also obviously Lipschitz in both the d_L and d_R metrics, and therefore any hyperbolic group is guaranteed a rich family of such functions.

6.3.2. Combable functions. We now introduce the class of *combable functions* on a hyperbolic group G.

Definition 6.31. Let G be word-hyperbolic with finite symmetric generating set S, and let $L \subset S^*$ be a combing of G with respect to S. A function $\phi : G \to \mathbb{Z}$ is *weakly combable* with respect to S, L (or *weakly combable* if S, L are understood) if there is a digraph Γ parameterizing L and a function $d\phi$ from the vertices of Γ to $\mathbb{Z}$, such that for any word $w \in L$ there is an equality

$$\phi(\mathsf{e}(w)) = \sum_i d\phi(\gamma_i(w))$$

(here $\mathsf{e}(w)$ on the left denotes an element of G and $\gamma_i(w)$ on the right denotes the vertices in Γ of the path corresponding to $w \in L$). If the maps e and γ are understood, by abuse of notation we write this formula as

$$\phi(w) = \sum_i d\phi(w_i)$$

A function ϕ is *combable* if it is weakly combable and is Lipschitz as a map from $G, d_L \to \mathbb{Z}$. It is *bicombable* if it is weakly combable and is Lipschitz both as a map from $G, d_L \to \mathbb{Z}$ and from $G, d_R \to \mathbb{Z}$.

A weakly combable function is *ergodic* (resp. almost ergodic) if there is an automaton Γ parameterizing L which has a unique maximal recurrent subgraph (resp. with maximal eigenvalue), and is *regular* if it is ergodic, and its recurrent subgraph is aperiodic.

WARNING 6.32. Remember that a combing L with respect to S can be parameterized by many different graphs Γ. If ϕ is weakly combable with respect to S, L then there is *some* digraph Γ parameterizing L for which $d\phi$ is a function on Γ. The particular parameterizing digraph Γ may definitely depend on ϕ.

REMARK 6.33. There is no strict logical necessity to restrict attention to functions with values in $\mathbb{Z}$. One can vary the definition and for any finitely generated group H define weakly combable H-functions, by defining $d\phi : \Gamma \to H$ and replacing sum by group multiplication in H. Since H is finitely generated, it makes sense to talk about left Lipschitz and right Lipschitz functions from G to H and therefore to define combable and bicombable H-functions. Notice with this definition that any homomorphism $G \to H$ is a bicombable H-function.

EXAMPLE 6.34. Word length is bicombable.

REMARK 6.35. Theorem 6.29 remains true, and with essentially the same proof, when S is an *asymmetric* generating set which generates G as a semigroup. For semigroup generators, one must slightly change the definition of a combing to say that words in L represent shortest *directed* paths to their endpoints, rather than geodesics in $C_S(G)$. It follows that Example 6.34 remains true in the more general context of word length with respect to an asymmetric set of generators for G (as a semigroup).

The definition of weakly combable depends quite strongly on the choice of the generating set S, as the following example shows.

EXAMPLE 6.36. Let $G = \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$, and let the factors be generated by a and b respectively. Define $f : G \to \mathbb{Z}$ by

$$f(w) = \begin{cases} n \text{ if } w = a^n \text{ for some } n \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

Then f is weakly combable with respect to the generating set a, a^{-1}, b; a digraph to calculate f is depicted in Figure 6.5. On the other hand, f is not weakly combable

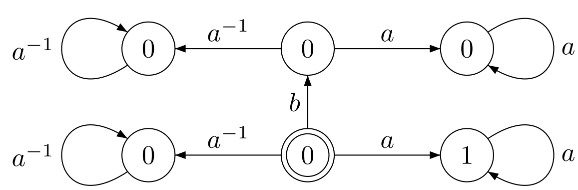


FIGURE 6.5. A digraph to calculate f

with respect to the generating set $ab, a^{-1}b, b$. Note for this generating set that $(ab)^n$ is the unique geodesic representing its value in G, and therefore $(ab)^n \in L$ for all n. Suppose to the contrary that f is weakly combable with respect to $ab, a^{-1}b, b$, so there is a finite digraph Γ parameterizing L and a function $df : \Gamma \to \mathbb{Z}$ as in Definition 6.31. Since Γ is finite, there is a constant C such that $|f(w) - f(ws)| \leq C$ whenever w and ws are words in L which differ by right multiplication by a single generator. Yet $f((ab)^{2n}) = 2n$ for $n \geq 0$, whereas $f((ab)^{2n+1}) = 0$, so no such pair L, Γ can exist.

Example 6.36 shows that the property of weak combability is contingent, and perhaps not so useful. By contrast, the Independence Theorem (Theorem 6.39, to be proved shortly) shows that combability is *independent* of the choice of generating set. For this reason, combable functions are much more useful and interesting than weakly combable functions.

We introduce some definitions which will be useful in what follows.

DEFINITION 6.37. Let G be hyperbolic with finite symmetric generating set S. Let L be a combing with respect to S. Let B be the ball of radius N about id in G with the metric inherited from $C_S(G)$, and let Σ be a finite set. A *tile set* is a map

$$T : B \times G \to \Sigma$$

such that for any pair of words $w, ws \in L$ where $s \in S$, the map $T(\cdot, \mathsf{e}(ws)) : B \to \Sigma$ depends only on $T(\cdot, \mathsf{e}(w))$ and s.

DEFINITION 6.38. Let T be a tile set, and let Γ be a digraph parameterizing L. The *fiber product* is the digraph Γ_T parameterizing L defined as follows. The vertices of Γ_T are the functions of the form

$$(T(\cdot, \mathsf{e}(w)), \gamma(w)) : B \to \Sigma \times \Gamma$$

and $(T(\cdot, \mathsf{e}(w)), \gamma(w))$ is joined to $(T(\cdot, \mathsf{e}(ws)), \gamma(ws))$ by an edge labeled s whenever $w, ws \in L$.

Geometrically, Γ_T can be thought of as a bundle over Γ whose fiber at each vertex v is the (finite) set of functions of the form $T(\cdot, \mathsf{e}(w)) : B \to \Sigma$ for all w satisfying $\gamma(w) = v$.

THEOREM 6.39 (Independence of combability). *Let $\phi : G \to \mathbb{Z}$ be combable with respect to some S', L'. Then for any other generating set S and any combing L with respect to S, the function ϕ is combable with respect to S, L.*

PROOF. If S, S' are two generating sets, and L, L' are two bijective geodesic combings, then every word in L' is quasigeodesic in $C_S(G)$ and by the Morse Lemma (Theorem 3.30, bullet (1)), (asynchronously) fellow travels the word in L with the same evaluation. That is, there are constants N and k such that the following is true:

(1) For all words w' in L' and w in L with $\mathsf{e}(w') = \mathsf{e}(w)$, the path w' (i.e. the set of $\mathsf{e}_i(w')$) is contained in the N neighborhood of the path w (i.e. the set of $\mathsf{e}_i(w)$) in $C_S(G)$. Furthermore, the path w' intersects the N neighborhood of *every* vertex on w (i.e. it comes uniformly close to every vertex on w)
(2) If $\mathsf{e}_i(w') \in B_N(\mathsf{e}_j(w))$ and $\mathsf{e}_l(w') \in B_N(\mathsf{e}_{j+1}(w))$, then $|l - i| < k$

Bullet (1) may be restated informally as saying that for every $w' \in L'$ and $w \in L$ with $\mathsf{e}(w') = \mathsf{e}(w)$, the path corresponding to w' is obtained by concatenating paths x_i of uniformly bounded length, whose endpoints are within a bounded distance of successive vertices of w.

Now, suppose ϕ is combable with respect to L'. Let Γ' be a digraph which parameterizes words in L' for which $d\phi : \Gamma' \to \mathbb{Z}$ is defined. Let B denote the ball of radius N around id in G with the metric inherited from $C_S(G)$.

We define a tile set T taking values in a certain finite set as follows. For each $g \in G$ and $h \in B$, let $w \in L$ and $w' \in L'$ evaluate to g and gh. That is, $\mathsf{e}(w) = g$

and $\mathsf{e}(w') = gh$. If some $\mathsf{e}_i(w')$ is not contained in the N neighborhood of any $\mathsf{e}_j(w)$, or if the N neighborhood of some $\mathsf{e}_j(w)$ does not intersect w' (i.e. if the conditions of bullet (1) above are violated), then $T(h,g) = \mathtt{E}$, an "out of range" symbol. Otherwise set

$$T(h,g) = (\phi(gh) - \phi(g), \gamma(w'))$$

in other words, the tuple consisting of the difference of ϕ on gh and g, and the vertex of Γ' corresponding to the endpoint of the path w'.

In words, for a fixed $g \in G$, the set of pairs h, g parameterizes the ball of radius N about g. For every element gh of this ball, there is a unique path in L' which evaluates to gh. If this path does not stay in the N neighborhood of the path in L evaluating to g, the value of T is out of range. Otherwise, T calculates the value of ϕ on the element gh (normalized by subtracting the value of ϕ on g) and the vertex of Γ' associated to the word of L' corresponding to gh.

Since ϕ is Lipschitz in the $C_{S'}(G)$ metric, it is also Lipschitz in the $C_S(G)$ metric, so the normalized values of ϕ on $B_N(g)$ are uniformly bounded, independent of $g \in G$. This shows that T takes values in a *finite* set. This is the only place where combability (as distinct to weak combability) is used in the proof. We will show that T is a *tile set*.

Remark 6.40. In fact, the second factor of T is by itself already a tile set; on a first reading, it is worth verifying this fact alone, and then seeing how it can be used to deduce the stronger claim about T.

To verify that T is a tile set, we just need to check that if $w, ws \in L$ then $T(\cdot, \mathsf{e}(ws))$ depends only on $T(\cdot, \mathsf{e}(w))$ and on s.

Let $h \in B$, and suppose $w' \in L'$ is such that $\mathsf{e}(w') = \mathsf{e}(ws)h \in G$. If the path w' is contained in the N neighborhood of the path ws, there is a factorization $w' = v'x$ in L' where $\mathsf{e}(v')$ is within distance N of $\mathsf{e}(w)$, and where x is a path in Γ' of length $\le k$. So for each $f \in B$ with $\mathsf{e}(v') = \mathsf{e}(w)f$ we can enumerate the set of all paths α in Γ' of length $\le k$ starting at $\gamma(v')$, and see whether $f\mathsf{e}(\alpha) = \mathsf{e}(s)h$. If no such f, α exists, then $T(h, \mathsf{e}(ws)) = \mathtt{E}$. Otherwise, the state $\gamma(w')$ can be deduced from the state $\gamma(v')$ and from x (this shows that the second factor of T is a tile set), and we can calculate

$$\phi(\mathsf{e}(ws)h) - \phi(\mathsf{e}(w)f) = \sum_i d\phi(\alpha_i)$$

If $h = \mathrm{id}$ then some such f, α is guaranteed to exist, by the discussion above. Hence $\phi(\mathsf{e}(ws)) - \phi(\mathsf{e}(w)f)$ can be calculated, and therefore for any $h \in B$ we can calculate $\phi(\mathsf{e}(ws)h) - \phi(\mathsf{e}(ws))$ without using w, and therefore $T(\cdot, \mathsf{e}(ws))$ depends only on $T(\cdot, \mathsf{e}(w))$ and on s, not on ws. This shows that T is a tile set.

If Γ is a digraph parameterizing L, we build the fiber product Γ_T. Since $\phi(\mathsf{e}(ws)) - \phi(\mathsf{e}(w)f)$ and $\phi(\mathsf{e}(w)f) - \phi(\mathsf{e}(w))$ depend only on $T(\cdot, w)$ and s, the value of $\phi(\mathsf{e}(ws)) - \phi(\mathsf{e}(w))$ depends only on s and the vertex $\gamma(w)$ of Γ_T. So we can define $d\phi$ as a function on the resolution $(\Gamma_T)_1$ of Γ_T, where the value of $d\phi$ on the vertex of $(\Gamma_T)_1$ corresponding to the edge from $\gamma(w)$ to $\gamma(ws)$ is equal to $\phi(\mathsf{e}(ws)) - \phi(\mathsf{e}(w))$.

By construction, $d\phi$ satisfies

$$\phi(\mathsf{e}(w)) = \sum_i d\phi(\gamma_i(w))$$

and therefore ϕ is combable with respect to S, L. □

NOTATION 6.41. Denote the class of combable and bicombable functions on G by $\mathfrak{C}(G)$ and $\mathfrak{B}(G)$ respectively.

LEMMA 6.42. *$\mathfrak{C}(G)$ and $\mathfrak{B}(G)$ are free Abelian groups.*

PROOF. If ϕ is (bi-)combable, then obviously so is $-\phi$.

Let ϕ_1, ϕ_2 be combable. Then they are combable with respect to some fixed combing S, L. Let Γ_1, Γ_2 be digraphs parameterizing L for which $d\phi_i : \Gamma_i \to \mathbb{Z}$ is defined. Define a new digraph Γ with one vertex for each pair of vertices from Γ_1, Γ_2 and with an edge labeled s from (v_1, v_2) to (v_1', v_2') if and only if there is an edge of Γ_i from v_i to v_i' labeled s for $i = 1, 2$. The initial vertex of Γ is the pair consisting of the initial vertices of Γ_1, Γ_2 respectively. Let Γ' be the subgraph of Γ consisting of the union of all directed paths starting at the initial vertex. Then Γ' parameterizes L, and $d(\phi_1 + \phi_2)$ is a function on Γ' defined by

$$d(\phi_1 + \phi_2)(v_1, v_2) = d\phi_1(v_1) + d\phi_2(v_2)$$

and therefore $\phi_1 + \phi_2$ is weakly combable. A sum of two functions which are Lipschitz in the d_L (resp. d_R) metric is Lipschitz in the d_L (resp. d_R) metric, so $\phi_1 + \phi_2$ is (bi-)combable if both ϕ_i are.

This shows that $\mathfrak{C}(G)$ and $\mathfrak{B}(G)$ are Abelian groups. Since they take values in $\mathbb{Z}$, they are torsion-free, and not infinitely divisible. □

EXAMPLE 6.43. Let $G = F_2 = \langle a, b \rangle$ and let $f : G \to \mathbb{Z}$ be defined by

$$f(w) = \begin{cases} |w| \text{ if } w \text{ starts with } a \\ 0 \text{ otherwise} \end{cases}$$

Then f is weakly combable; a digraph to calculate f is illustrated in Figure 6.6.

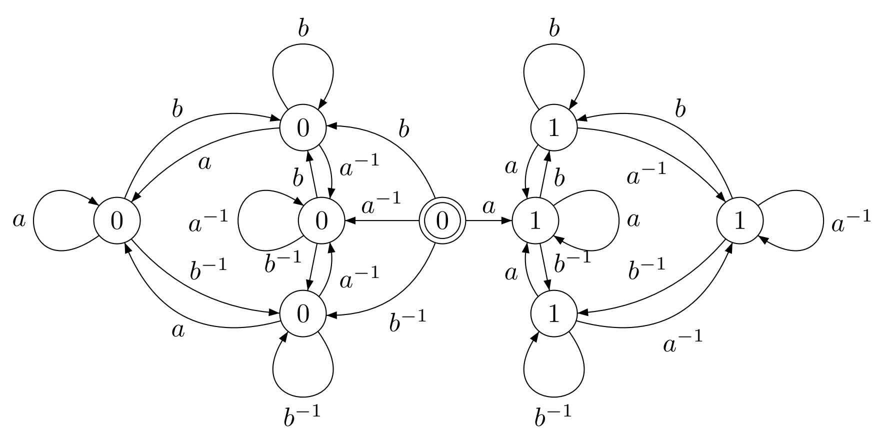


FIGURE 6.6. A digraph to calculate f

Moreover, f is Lipschitz in the d_L metric and therefore combable. However, $f(a^n) = n$ whereas $f(ba^n) = 0$ so f is not Lipschitz in the d_R metric, and is not bicombable.

6.3.3. Quasimorphisms. There are several natural operations which can be defined on functions $\phi : G \to \mathbb{R}$, including the following:

(1) The *adjoint* of ϕ, denoted ϕ^*, defined by

$$\phi^*(a) = \phi(a^{-1})$$

(2) The *antisymmetrization* of ϕ, denoted ϕ', defined by

$$\phi'(a) = \frac{1}{2}(\phi(a) - \phi(a^{-1})) = \frac{1}{2}(\phi - \phi^*)(a)$$

In general, neither operation preserves weak combability, although if both ϕ and ϕ^* are weakly combable, so is $2\phi'$.

LEMMA 6.44. *Suppose ϕ is weakly combable, and Lipschitz in the d_R metric. Then there is a constant C so that if $w \in L$ is expressed as a product of subwords $w = uv$ then $|\phi(w) - \phi(u) - \phi(v)| \le C$.*

PROOF. Let Γ be a digraph which parameterizes L. Let $u' \in L$ be any word such that $\gamma(u) = \gamma(u')$. Then $\phi(w) = \phi(u) + \phi(u'v) - \phi(u')$. Choose u' so that $|u'| \le |\Gamma|$. Since ϕ is Lipschitz in the d_R metric, there is a constant C_1 so that $|\phi(u'v) - \phi(v)| \le C_1$. Since $|u'|$ is bounded, there is a constant C_2 so that $|\phi(u')| \le C_2$. Hence

$$|\phi(w) - \phi(u) - \phi(v)| \le |\phi(u'v) - \phi(u') - \phi(v)| \le C_1 + C_2$$

proving the Lemma. □

In words, Lemma 6.44 says that ϕ is almost additive under decomposition.

LEMMA 6.45. *Suppose ϕ is bicombable. Then there is a constant C so that if $w \in L$ is expressed as a product of subwords $w = uv$ then*

$$|\phi^*(w) - \phi^*(u) - \phi^*(v)| \le C$$

PROOF. We have $w^{-1} = v^{-1}u^{-1}$ in G but not necessarily in L. Let $z \in L$ represent w^{-1}, and express z as a product of subwords $z = xy$ where $d_L(v^{-1}, x) \le \delta$ and $d_R(u^{-1}, y) \le \delta$. By Lemma 6.44, $|\phi(z) - \phi(x) - \phi(y)| \le C$. But $\phi(z) = \phi^*(w)$ whereas $|\phi^*(v) - \phi(x)| \le \delta C_1$ and $|\phi^*(u) - \phi(y)| \le \delta C_1$ for some C_1 because ϕ is bicombable (and therefore Lipschitz in both d_L and d_R). The Lemma now follows from the triangle inequality. □

THEOREM 6.46. *Let $\phi : G \to \mathbb{Z}$ be bicombable. Then the antisymmetrization ϕ' is a quasimorphism.*

PROOF. Let $u, v \in L$ be arbitrary, and let $w \in L$ satisfy $\mathsf{e}(w) = \mathsf{e}(u)\mathsf{e}(v)$. Then we can write $u = u'x$, $v = yv'$ and $w = w_1w_2$ as words in L so that $d_L(y, x^{-1}) \le \delta$, $d_L(u, w_1) \le \delta$ and $d_R(v, w_2) \le \delta$, by δ-thinness of triangles in $C_S(G)$.

Now apply Lemma 6.44, Lemma 6.45 and antisymmetry. □

REMARK 6.47. Say that a function $\phi : G \to \mathbb{R}$ is *almost antisymmetric* if there is a constant C so that $|\phi(a) + \phi^*(a)| \le C$ for all $a \in G$. The arguments above can be modified to show that an almost antisymmetric bicombable function is a quasimorphism.

Theorem 6.46 can be used to give a surprisingly simple construction of non-trivial quasimorphisms on any hyperbolic group.

EXAMPLE 6.48. Let G be hyperbolic, and let T be a finite *asymmetric* set which generates G as a semigroup. Let $w_T : G \to \mathbb{Z}$ be word length with respect to T. Then define

$$h_T(a) = w_T(a) - w_T(a^{-1})$$

for all $a \in G$.

By Remark 6.35, w_T is bicombable and therefore h_T is a quasimorphism.

In fact, it is straightforward to give a direct proof that h_T is a quasimorphism,. Let S be the symmetrization of T, and construct the Cayley graph $C_S(G)$. First of all, it is obvious that h_T is Lipschitz in both the d_R and the d_L metrics.

Secondly, every word in T is a path in $C_S(G)$ (but not conversely). A shortest path in $C_S(G)$ from id to a representing a word in T will be called a *realizing path* for a. Since every element in S can be written as a word of bounded length in T, there are uniform constants k, ϵ so that realizing paths are k, ϵ quasigeodesic in $C_S(G)$. In particular, if l_a and $l_{a^{-1}}$ are realizing paths for a and a^{-1} respectively, then l_a and $al_{a^{-1}}$ are δ' close for some δ' not depending on a. So if u is arbitrary, and $u = vw$ where v is on a realizing path for u, then w^{-1} is within distance δ' of a realizing path for u^{-1}. It follows that there is a constant C such that

$$|h_T(u) - h_T(v) - h_T(w)| \le C$$

for any such factorization. In other words, h_T is almost additive under decomposition.

Now, if a, b are arbitrary, and l_a, l_b, l_{ab} are realizing paths for a, b, ab respectively, then l_a, al_b, l_{ab} are three sides of a δ' thin quasigeodesic triangle. This triangle can be decomposed into six segments which are δ' close in pairs. Since h_T is antisymmetric, and Lipschitz in both d_L and d_R, the values of h_T on paired segments almost cancel. Since h_T is almost additive under decomposition, $h_T(ab)$ and $h_T(a) + h_T(b)$ are almost equal, and we are done. This shows that h_T is a quasimorphism.

For typical asymmetric T, the function h_T is unbounded. This is not completely trivial, but follows from estimates on the length of anti-aligned translates of an axis (compare with Remark 3.12). When G is nonelementary, by varying the choice of generating sets T and taking infinite (L_1) linear combinations, one can construct a subspace of $Q(G)$ with dimension $2^{\aleph_0}$, giving a new proof of the main theorem of Epstein–Fujiwara ([**78**], Thm. 1.1).

6.4. Counting quasimorphisms

6.4.1. Greedy algorithm. In § 3.5 we discussed Fujiwara's construction of counting quasimorphisms associated to an action of a group G on a δ-hyperbolic graph X. In the special case that G is a hyperbolic group, and X is the Cayley graph of G with respect to a finite generating set S, such quasimorphisms were constructed first by Epstein–Fujiwara [**78**], generalizing Brooks [**27**]. Our aim in this section and the next is to show that counting quasimorphisms are bicombable.

For the sake of clarity, we spell out the definition of Epstein–Fujiwara counting quasimorphisms.

DEFINITION 6.49. Let G be a hyperbolic group with symmetric generating set S. Let σ be an oriented simplicial path in the Cayley graph $C_S(G)$ and let σ^{-1} denote the same path with the opposite orientation. For γ an oriented simplicial path in $C_S(G)$, let $|\gamma|_\sigma$ denote the maximal number of disjoint copies of σ contained in γ. For $a \in G$, define

$$c_\sigma(a) = \text{dist}(\text{id}, a) - \inf_\gamma(\text{length}(\gamma) - |\gamma|_\sigma)$$

where the infimum is taken over *all* directed paths γ in $C_S(G)$ from id to a.

Define a (small) *counting quasimorphism* to be a function of the form

$$h_\sigma(a) := c_\sigma(a) - c_{\sigma^{-1}}(a)$$

This is a special case of Fujiwara's construction in § 3.5 and therefore when $|\sigma| \geq 2$, Lemma 3.46 applies, and realizing paths are (uniformly) quasigeodesic in $C_S(G)$.

Let σ be a string. If w is a word, let $|w|_\sigma$ count the maximal number of disjoint copies of σ in w. Similarly, let $|w|'_\sigma$ count disjoint copies of σ in w using the *greedy algorithm*. In other words, define $|w|'_\sigma$ inductively on the length of w by the equality

$$|w|'_\sigma = |v|'_\sigma + 1$$

where v is the word obtained from w by deleting the prefix up to and including the first occurrence of σ in w.

The advantage of $|\cdot|'_\sigma$ over $|\cdot|_\sigma$ is that it is evident from the definition that $d|\cdot|'$ can be calculated by a finite state automaton. On the other hand, we have the following:

LEMMA 6.50 (Greedy is good). *The functions $|\cdot|_\sigma$ and $|\cdot|'_\sigma$ are equal.*

PROOF. Suppose not, and let w be a shortest word such that $|w|_\sigma$ and $|w|'_\sigma$ are not equal. By definition, $|w|'_\sigma < |w|_\sigma$, and since w is the shortest word with this property, by comparing the values of the two functions on prefixes of w, we conclude $|w|'_\sigma = |w|_\sigma - 1$. Since w is the shortest word with this property, the suffix of w must be a copy of σ that is counted by $|\cdot|_\sigma$ but not by $|\cdot|'_\sigma$. Hence the greedy algorithm must count a copy of σ that overlaps this suffix. Deleting the terminal copy of σ reduces the values of both $|\cdot|_\sigma$ and $|\cdot|'_\sigma$ by 1, contrary to the hypothesis that w was shortest. □

6.4.2. Counting quasimorphisms are bicombable.

THEOREM 6.51 (Calegari–Fujiwara [**50**]). *Let G be hyperbolic, and let h_σ be an Epstein–Fujiwara counting quasimorphism. Then h_σ is bicombable.*

PROOF. We give a somewhat informal proof, which can be made rigorous by translating it into the language of tile sets, and following the model of Theorem 6.39.

Fix a hyperbolic group G and a symmetric generating set S. Let L be a combing for G. Remember that this means that L is a prefix-closed regular language of geodesics in G (with respect to the fixed generating set S) for which the evaluation map is a bijection $L \to G$. If $w \in L$ corresponds to $\overline{w}$ in G, let γ_w be the path in $C_S(G)$ from id to $\overline{w}$.

Let σ be a string. We will show that both c_σ and $c_{\sigma^{-1}}$ are weakly combable with respect to the generating set S and any combing L. By bullet (2) of Lemma 3.45, these functions are Lipschitz in the d_L metric, and therefore combable. Lemma 6.42 implies that their difference h_σ is also combable; since it is a quasimorphism, it is bicombable.

In the remainder of the proof, for the sake of clarity, we abbreviate c_σ to c.

Fix a word $w \in L$. By Lemma 3.46, a realizing path α for $\overline{w}$ is a K, ϵ quasigeodesic, and therefore by the Morse Lemma, there is a constant N depending only on δ, K, ϵ (and not on w) so that α and γ_w are contained in N-neighborhoods of each other. Hence every vertex of α is contained in the N-neighborhood of some vertex of γ_w and conversely. For each i, let $B_N(\gamma_w(i))$ denote the N-neighborhood of $\gamma_w(i)$. By uniform quasigeodesity of α, and geodesity of γ, if $p \in B_N(\gamma_w(i))$ and $q \in B_N(\gamma_w(i+1))$ are both on α, then the segment of α from p to q has uniformly bounded length. Let $p \in B_N(\gamma_w(i))$ for some i. Say a path γ' from id to p is *admissible* if it is K, ϵ-quasigeodesic, and if for all $j < i$ the path γ' intersects $B_N(\gamma_w(j))$. Thus, an admissible path is obtained by concatenating paths of bounded length whose endpoints are contained in N-neighborhoods of successive vertices of γ_w.

For each $p \in B_N(\gamma_w(i))$ and each path γ' from id to p, recall that $|\gamma'|$ is the maximal number of disjoint copies of σ in γ'. By Lemma 6.50, the greedy algorithm picks out $|\gamma'|$ specific disjoint copies which we refer to as the *greedy copies* of σ in γ'; let $\sigma(\gamma')$ be the biggest prefix of σ which is a suffix of γ' and which is *disjoint* from the greedy copies of σ in γ'. Let X denote the set of possible values of $\sigma(\gamma')$. Note that $|X| = |\sigma|$, since the values of X are in bijection with proper prefixes of σ. One can think of the set X as the states of an automaton that reads a word, and finds the greedy copies of σ in that word.

We define a function T as follows. The domain of T is $B_N(\text{id}) \times X \times G$. Fix $h \in B_N(\text{id})$ and $\gamma_w(i) \in G$. Let $g = \gamma_w(i)h \in B_N(\gamma_w(i))$. For each $x \in X$, consider the set of all admissible paths γ' from id to g that satisfy $\sigma(\gamma') = x$. If no such path exists, define $T(h, x, \gamma_w(i)) = \mathtt{E}$, an "out of range" symbol. Otherwise, define

$$c(g, x) = \text{dist}(\text{id}, g) - \inf_{\gamma'}(\text{length}(\gamma') - |\gamma'|)$$

where the infimum is taken over γ' as above. Notice that $\max_x c(g, x) = c(g)$ if there is some admissible realizing path. In particular, $\max_x c(\gamma_w(i), x) = c(\gamma_w(i))$. If some γ' exists as above, define

$$T(h, x, \gamma_w(i)) = c(\gamma_w(i)) - c(g, x)$$

If there is any admissible path γ' from id to g that ends in state x, there is such a path obtained by composing a realizing path for $\gamma_w(i)$ with a suffix of bounded length. Together with bullet (2) of Lemma 3.45, this implies that T takes values in a finite set.

Suppose we know the value of T on $B_N(\text{id}) \times X \times \gamma_w(i)$. Let $h \in B_N(\text{id})$, and define $g' = \gamma_w(i+1)h$. Any admissible path from id to g' is obtained by concatenating an admissible path from id to some $g \in B_N(\gamma_w(i))$ with a path of bounded length. So if we can compute $d(\text{id}, g') - d(\text{id}, g)$ we can compute $c(g', x) - c(g, y)$ for any $x, y \in X$. Since G is hyperbolic, and γ_w is geodesic, we can keep track of relative distances from id to points in the ball of radius N about points on γ_w, and therefore we can compute $d(\text{id}, g') - d(\text{id}, g)$ by keeping track of only a finite amount of information at each stage. We define a digraph parameterizing L that keeps track at each stage of the following two pieces of information, thought of as functions on the ball of radius N about the current vertex in $C_S(G)$:

(1) The relative distances from id
(2) The value of T

By the discussion above, this is a finite digraph, and dc is well-defined as a function on the vertices of its first refinement (cf. Theorem 6.39). Hence c_σ and $c_{\sigma^{-1}}$ are combable, and the proof follows. □

REMARK 6.52. The pair consisting of T and relative distance to id is almost a tile set, except that the domain is slightly larger (since T depends, in addition to $B_N(\text{id})$ and G, on the choice of an element in the finite set X). Otherwise, the proof is conceptually very similar to that of Theorem 6.39.

6.5. Patterson–Sullivan measures

The crucial difficulty in extending Theorem 6.13 to general word-hyperbolic groups is the fact that the digraphs associated to arbitrary word-hyperbolic groups are (typically) not recurrent. This means that the stationary Markov chains obtained by generalizing the construction of § 6.1.4 are not typically ergodic, and the Perron–Frobenius Theorem (i.e. Theorem 6.6) does not directly apply.

The first important result we use in this section is Coornaert's Theorem, which says that in a non-elementary word-hyperbolic group G, if we fix a finite generating set, there are constants $\lambda > 1$ and $K \geq 1$ so that the number of words of length n is bounded between $K^{-1}\lambda^n$ and $K\lambda^n$ for all n. This implies that one can find a digraph parameterizing a combing of G which is *almost semisimple* — that is, the eigenspace of largest absolute value is *diagonalizable*, and the system (measurably) decomposes into a finite number of independent ergodic subsystems. Consequently, most long geodesics in G can be partitioned into finitely many families, each (more-or-less) parameterized by random walks on a *recurrent* digraph whose associated stationary Markov chain is ergodic, and obeys a central limit theorem.

A priori, there is no apparent way to compare long geodesics in different families. However, in place of recurrence of a single digraph, one can use the ergodicity of the action of G *at infinity*, on the boundary ∂G with its Patterson–Sullivan measure. A typical infinite geodesic in one family can be translated by left-multiplication to within bounded distance of a typical infinite geodesic in any other family. A bicombable function is almost invariant under both left and right multiplication by elements of bounded size, so the distribution of values on a typical infinite geodesic in one family is *the same* as the distribution on a typical infinite geodesic in the other. In other words, the values of the function on typical paths in one family have the same distribution as the values of typical paths in any other, and we obtain a central limit theorem for the group as a whole. The next few sections flesh out the details of this scheme.

6.5.1. Some linear algebra.

Let Γ be a finite pointed digraph. Let V be the real vector space spanned by the vertices of Γ, and let $\langle \cdot, \cdot \rangle$ be the inner product on V for which the vertices are an orthonormal basis.

The vertices of Γ are denoted v_i for $i \in \{1, \cdots, n\}$. We let v_1 denote the initial vertex. For a vector $v \in V$, let $|v|$ denote the L^1 norm of v. That is,

$$|v| = \sum_i |\langle v, v_i \rangle|$$

For brevity, let $\mathbf{1}$ denote the vector with all co-ordinates equal to 1, so for a non-negative vector v, there is equality $|v| = \langle v, \mathbf{1} \rangle$.

The digraphs Γ that parameterize combings of hyperbolic groups are not completely general, but satisfy a number of special properties. We formalize these properties as follows. Let M denote the adjacency matrix of Γ, so that the number of directed paths in Γ of length n from v_i to v_j is

$$(v_i)^T M^n v_j = \langle v_i, M^n v_j \rangle = (M^n)_{ij}$$

DEFINITION 6.53. A digraph Γ is *almost semisimple* if it satisfies the following properties.

(1) There is an *initial vertex* v_1
(2) For every $i \neq 1$ there is a directed path in Γ from v_1 to v_i
(3) There are constants $\lambda > 1, K \geq 1$ so that

$$K^{-1}\lambda^n \leq |v_1^T M^n| \leq K\lambda^n$$

for all positive integers n

In what follows we will assume that Γ is almost semisimple.

LEMMA 6.54. *Suppose Γ is almost semisimple. Then λ is the largest real eigenvalue of M. Moreover, for every eigenvalue ξ of M either $|\xi| < \lambda$ or else the geometric and the algebraic multiplicities of ξ are equal.*

PROOF. It is convenient to work with M^T in place of M. To prove the lemma, it suffices to prove analogous facts about the matrix M^T. Corresponding to the Jordan decomposition of M^T over $\mathbb{C}$, let $\xi_1, \ldots, \xi_m$ be the eigenvalues of the corresponding Jordan blocks (listed with multiplicity).

Bullet (2) from Definition 6.53 implies that for any v_i, there is an inequality $|(M^n)^T v_i| \leq C_i |(M^n)^T v_1|$ for some constant C_i. Since the v_i span V, and since V is finite dimensional, there is a constant C such that for all $w \in V$ there is an inequality $|(M^n)^T w| \leq C|(M^n)^T v_1||w|$.

For each i, there is some w_i in the ξ_i-eigenspace for which

$$|(M^n)^T w_i| \geq \text{constant} \cdot n^{k-1}|\xi_i|^n$$

where k is the dimension of the Jordan block associated to ξ_i. Since $|(M^n)^T w_i| \leq C|(M^n)^T v_1||w_i|$, by bullet (3) from Definition 6.53, either $|\xi| < \lambda$ or $|\xi| = \lambda$ and $k = 1$.

By the Perron–Frobenius theorem for non-negative matrices, M^T has a largest real eigenvalue λ' such that $|\xi| \leq \lambda'$ for all eigenvalues ξ. We must have $\lambda' = \lambda$ by the estimates above. Note that M has the same spectrum as M^T with the same multiplicity, and that all the ξ eigenspaces of M are diagonalizable for $|\xi| = \lambda$. □

For any vector $v \in V$, decompose $v = \sum_\xi v(\xi)$ into the components in the generalized eigenspaces of the eigenvalues ξ. Since any two norms on $V \otimes \mathbb{C}$ are equivalent, there is a constant $K > 1$ such that

$$K^{-1} \leq \frac{|M^n v|}{\sum_\xi |M^n v(\xi)|} \leq K$$

and similarly for M^T.

LEMMA 6.55. *For any vector $v \in V$, the following limit*

$$\rho(v) := \lim_{n\to\infty} n^{-1} \sum_{i \leq n} \lambda^{-i} M^i v$$

exists and is equal to $v(\lambda)$.

PROOF. We suppress v in the notation that follows. For each eigenvector ξ define

$$\rho_n(\xi) = n^{-1}\sum_{i\le n}\lambda^{-i}M^i v(\xi)$$

And set $\rho_n = \sum_\xi \rho_n(\xi)$. With this notation, $\rho = \lim_{n\to\infty}\rho_n$, and we want to show that this limit exists.

By Lemma 6.54, for each ξ, either $|\xi| < \lambda$ or $v(\xi)$ is a ξ-eigenvector. In the first case, $\rho_n(\xi)\to 0$. In the second case, either $\xi = \lambda$, or else the vectors $\lambda^{-i}M^i v(\xi)$ become equidistributed in the unit circle in the complex line of $V\otimes\mathbb{C}$ spanned by $v(\xi)$. It follows that $\rho_n(\xi)\to 0$ unless $\xi=\lambda$.

So $n^{-1}\sum_{i\le n}\lambda^{-i}M^i v(\xi)\to 0$ unless $\xi=\lambda$, in which case $\rho_n(\lambda) = v(\lambda)$ is constant. □

Since every eigenvalue of M with largest (absolute) value has geometric multiplicity equal to its algebraic multiplicity, the same is true of the transpose M^T. The same argument as Lemma 6.55 implies

LEMMA 6.56. *For any vector $v\in V$, the following limit*

$$\ell(v) := \lim_{n\to\infty} n^{-1}\sum_{i\le n}\lambda^{-i}(M^T)^i v$$

exists, and $(\ell(v))^T$ is the projection of v^T onto the left λ eigenspace of M.

For any v_i, the partial sums $\rho_n(v_i)$ are non-negative real vectors so if v is non-negative, so is $\rho(v)$. Similarly, if v is non-negative, so is $\ell(v)$.

PROPOSITION 6.57. *For any $v, w\in V$ there is equality*

$$\langle \ell(v), w\rangle = \langle \ell(v), \rho(w)\rangle = \langle v, \rho(w)\rangle$$

PROOF. By definition,

$$\begin{aligned}
\langle \ell(v), \rho(w)\rangle &= \lim_{n\to\infty}\left(n^{-1}\sum_{i\le n}\lambda^{-i}v^T M^i\right)\left(n^{-1}\sum_{j\le n}\lambda^{-j}M^j w\right)\\
&= \lim_{n\to\infty} n^{-2}\sum_{i,j\le n}\lambda^{-i-j}v^T M^{i+j}w\\
&= \lim_{n\to\infty} n^{-2}\sum_{k\le 2n}(n+1-|n-k|)\lambda^{-k}v^T M^k w\\
&= \lim_{n\to\infty} n^{-2}\sum_{k\le 2n}(n+1-|n-k|)\lambda^{-k}\ell(v)^T M^k w\\
&= \lim_{n\to\infty} n^{-2}n(n+1)\ell(v)^T w = \langle \ell(v), w\rangle
\end{aligned}$$

where the third last equality follows from the "almost periodicity" of $\lambda^{-1}M$ so that all terms except the (left and right) λ-eigenvalues cancel over any long consecutive sequence of indices. We get $\langle \ell(v), \rho(w)\rangle = \langle v, \rho(w)\rangle$ by the same reason. □

Recall that a *component* of Γ is a maximal recurrent subgraph C; i.e. a subgraph with the property that there is a directed path from any vertex to any other vertex. Each component C has its own adjacency matrix, with biggest real eigenvalue $\xi(C)$. Since C is a subgraph of Γ, we must have $\xi(C)\le\xi(\Gamma)=\lambda$ for any C.

LEMMA 6.58. *Let Γ be almost semi-simple. If C, C' are distinct components with $\xi(C) = \xi(C') = \lambda$ then there is no directed path from C to C'.*

PROOF. Recall the Landau notation $f(x) = \Theta(g(x))$ if the ratio $f(x)/g(x)$ is bounded away from zero and away from infinity.

Let u be a vertex in C and v a vertex in C' such that there is a directed path γ from u to v. Since C is recurrent, Proposition 6.7 implies that there are $\Theta(\lambda^n)$ directed paths in C starting at u of length n, and similarly for paths in C' starting at v. There is a constant k so that each vertex in C can be joined by a path of length at most k to some v. So for each pair of integers $i, n-i$ consider the set of paths of length between n and $n+k$ which consist of an initial segment of length i in C starting at u, followed by a path of length $\leq k$ to v, followed by a terminal segment of length $n-i$ in C. The number of such paths for fixed i is $\Theta(\lambda^n)$, so the number of paths for varying i is $\Theta(n\lambda^n)$. But if Γ is almost semi-simple, the number of paths of length between n and $n+k$ (of any kind) is $\Theta(\lambda^n)$, so we obtain a contradiction. □

6.5.2. Coornaert's Theorem and Patterson–Sullivan measures. Let G be a non-elementary word-hyperbolic group with generating set S. For $g \in G$, let $|g|$ denote word length with respect to S.

DEFINITION 6.59. The *Poincaré series* of G is the series

$$\zeta_G(s) = \sum_{g \in G} e^{-s|g|}$$

This series diverges for all sufficiently small s, and converges for all sufficiently large s. The *critical exponent* is the supremum of the values of s for which the series diverges. Similar zeta functions appear in many contexts, for example in number theory and dynamics. The best results can be expected when the series diverges at the critical exponent.

THEOREM 6.60 (Coornaert, [**56**] Thm. 7.2). *Let G be a non-elementary word-hyperbolic group with generating set S. Let G_n be the set of elements of word length n. Then there are constants $\lambda > 1, K \geq 1$ so that*

$$K^{-1}\lambda^n \leq |G_n| \leq K\lambda^n$$

for all positive integers n.

It follows from Theorem 6.60 that the critical exponent of the Poincaré series is equal to $\log(\lambda)$, and the series $\zeta_G(\log(\lambda))$ diverges.

For each n, let ν_n be the probability measure on G defined by

$$\nu_n = \frac{\sum_{|g| \leq n} \lambda^{-|g|}\delta_g}{\sum_{|g| \leq n} \lambda^{-|g|}}$$

where δ_g is the Dirac measure on the element g. The measure ν_n extends trivially to a probability measure on the compact space $G \cup \partial G$, where ∂G denotes the ideal (Gromov) boundary of G.

DEFINITION 6.61. A weak limit $\nu := \lim_{n\to\infty} \nu_n$ is a *Patterson–Sullivan* measure associated to S.

Since the Poincaré series diverges at the critical exponent, the support of ν is contained in ∂G.

It is convenient, for the sake of computations, to work with a slightly different normalization of ν. For each n, let $\widehat{\nu}_n$ be the measure on G defined by

$$\widehat{\nu}_n = \frac{1}{n}\sum_{|g|\le n} \lambda^{-|g|}\delta_g$$

and let $\widehat{\nu} := \lim_{n\to\infty} \widehat{\nu}_n$ be a weak limit. Of course the measures $\widehat{\nu}_n$ and ν_n are proportional for each n. Moreover, by Theorem 6.60, the constant of proportionality is bounded above and below.

REMARK 6.62. In fact, the limit of the $\widehat{\nu}_n$ exists and is well-defined. This is guaranteed by an explicit formula for $\widehat{\nu}$, which is given in § 6.5.3.

The group G acts on itself by left-multiplication. This action extends continuously to a left action $G \times \partial G \to \partial G$. Patterson–Sullivan measures enjoy a number of useful properties, summarized in the following theorem.

THEOREM 6.63 (Coornaert, [**56**] Thm. 7.7). *Let ν be a Patterson–Sullivan measure. The action of G on ∂G preserves the measure class of ν. Moreover, the action of G on $(\partial G, \nu)$ is ergodic.*

The meaning of ergodicity for a group action which preserves a measure class but not a measure is that for any $A, B \subset \partial G$ with positive ν-measure, there is $g \in G$ with $\nu(gA \cap B) > 0$. Since ν and $\widehat{\nu}$ are proportional, the action of G on ∂G is also ergodic for the $\widehat{\nu}$ measure.

In fact, Coornaert proves the stronger fact that there is a constant $K > 1$ so that for any $s \in S$ there is an inequality

$$K^{-1} \le \frac{d(s_*\nu)}{d\nu} \le K$$

and the same is true for the measure $\widehat{\nu}$, though we do not use this stronger fact.

6.5.3. Construction of stationary measure. Throughout the sequel we fix the following notation.

Let G be word-hyperbolic, and $\phi : G \to \mathbb{Z}$ a bicombable function. Fix a finite generating set S, and let $L \subset S^*$ be a combing of G with respect to S. Since ϕ is bicombable, $d\phi$ exists as a map from $\Gamma \to \mathbb{Z}$ for some digraph Γ parameterizing L, by Theorem 6.39.

Let M denote the adjacency matrix of Γ, acting on V, the space of real-valued functions on the vertices of Γ. Let $v_1 \in V$ be the function taking the value 1 on the initial vertex, and 0 on all other vertices. Let $\mathbf{1}$ denote the constant function taking the value 1 on every vertex of Γ.

For each n let X_n denote the set of walks of length n on Γ (starting at an arbitrary vertex) and Y_n the set of walks of length n starting at the initial vertex. There are restriction maps $X_{n+1} \to X_n$ and $Y_{n+1} \to Y_n$ for each n, with inverse limits X and Y. Evaluation of words gives rise to bijections $Y_n \to G_n$ for all n; taking limits, there is a map $Y \to \partial G$, called the *endpoint map*, taking an infinite word to the endpoint of the corresponding geodesic ray in G.

LEMMA 6.64 (Coornaert–Papadopoulos). *The endpoint map $Y \to \partial G$ is surjective, and bounded-to-one.*

See [**57**] for a proof.

REMARK 6.65. In fact, obtaining a bound on the size of the preimage of a point in ∂G is straightforward. If γ, γ' are infinite geodesics corresponding to paths in Y with the same endpoint in ∂G, then their Hausdorff distance is bounded by δ, the constant of hyperbolicity of G. For any point $\gamma_i \in \gamma$, let B_i denote the ball of radius δ about γ_i. Then γ' must intersect B_i, and the prefix of γ' up to this point of intersection is uniquely determined by the fact that γ' corresponds to a path in Y. Hence γ' may be thought of as an element of the inverse limit of a partially defined system of maps $B_i \to B_{i-1}$. Since $|B_i| \le C$ for all i for some constant C, the cardinality of this inverse limit is also bounded by C.

Each $g \in G_n$ corresponds to a unique word $w \in L$ and a unique path $y \in Y_n$. For each $m > n$ the projection $p : Y_m \to Y_n$ determines a subset $p^{-1}(y) \in Y_m$ and a corresponding subset of G_m. The set of $h \in G$ corresponding to words z in some Y_m which restricts to a fixed y is called the *cone* of g, and denoted cone(g). Note that cone(g) depends on L, but not on Γ. For each fixed n, we can define a measure $\widehat{\nu}$ on G_n by

$$\widehat{\nu}(g) = \lim_{m\to\infty} \widehat{\nu}_m(\text{cone}(g))$$

(an explicit formula is given below). Identifying G_n with Y_n, we obtain a measure on Y_n for each n which by abuse of notation we denote $\widehat{\nu}$. Observe that these measures for different n have the following compatibility property: for each $y \in Y_n$ and each $m > n$, there is an equality $\widehat{\nu}(p^{-1}(y)) = \widehat{\nu}(y)$ where $p : Y_m \to Y_n$ is the restriction map. This compatibility property means that we can define a measure $\widehat{\nu}$ on Y by the formula

$$\widehat{\nu}(p^{-1}(y)) = \widehat{\nu}(y) = \lim_{n\to\infty} \widehat{\nu}_n(\text{cone}(g))$$

where $p : Y \to Y_n$ is restriction. Since the cylinders $p^{-1}(y)$ generate the Borel σ-algebra of Y, this defines a unique measure $\widehat{\nu}$ on Y which by construction pushes forward under $Y \to \partial G$ to the measure $\widehat{\nu}$ of the same name on ∂G.

We can obtain an explicit formula for the value of $\widehat{\nu}$ on an element $g \in G_n$ or the corresponding element $y \in Y_n$ or cylinder $p^{-1}(y) \subset Y$. By definition, for any $g \in G_n$ and any $m \ge n$ we have

$$\widehat{\nu}_m(\text{cone}(g)) = \frac{1}{m} \sum_{\substack{h\in\text{cone}(g)\\ |h|\le m}} \lambda^{-|h|}$$

Let $v_g \in \Gamma$ be the last vertex of y. Then we can rewrite this formula as

$$\widehat{\nu}_m(\text{cone}(g)) = \frac{1}{m}\lambda^{-n} \sum_{i\le m-n} \lambda^{-i}\langle (M^i)^T v_g, \mathbf{1}\rangle$$

and therefore by taking limits $m \to \infty$ we obtain the formula

$$\widehat{\nu}(\overline{\text{cone}(g)}) = \lambda^{-n}|\ell(v_g)| = \lambda^{-n}\langle \ell(v_g), \mathbf{1}\rangle = \lambda^{-n}\langle v_g, \rho(\mathbf{1})\rangle$$

where overline denotes closure in $G \cup \partial G$, and where we have used the property that $\ell(\cdot)$ of a non-negative vector is non-negative, and Proposition 6.57 for the last equality.

The measure $\widehat{\nu}$ on Y is typically not invariant under the shift map $S : X \to X$. In fact, $S(Y) \cap Y = \emptyset$ if the initial vertex has no incoming edges. We define a measure μ on X by

$$\mu := \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{n} S_*^i \widehat{\nu}$$

and observe that the result is manifestly invariant by S. Using the explicit formula for $\widehat{\nu}$ on Y_n and Y we can derive an explicit formula for μ, showing that μ is well-defined.

Let $v_j \in \Gamma = X_0$ be an arbitrary vertex. By abuse of notation, we let $p : X \to X_0$ denote the restriction map of an infinite path to its initial vertex (this is similar to, but should not be confused with, the restriction maps $p : Y_m \to Y_n$ discussed earlier). We will calculate $\mu(p^{-1}(v_j))$. For each n we can calculate

$$S_*^n \widehat{\nu}(p^{-1}(v_j)) = \lambda^{-n} \sum_{\substack{y\in Y_n \\ S^n y = v_j}} \langle v_j, \rho(\mathbf{1})\rangle$$

On the other hand, the number of $y \in Y_n$ with $S^n y = v_j$ is exactly equal to the number of directed paths in Γ of length n which end at v_j, which is $\langle v_1, M^n v_j\rangle$. It follows that

$$\begin{aligned}
\mu(p^{-1}(v_j)) &= \lim_{n\to\infty} \langle v_j, \rho(\mathbf{1})\rangle \langle v_1, \frac{1}{n}\sum_{i\le n} \lambda^{-n} M^n v_j\rangle \\
&= \langle v_j, \rho(\mathbf{1})\rangle\langle v_1, \rho(v_j)\rangle = \langle v_j, \rho(\mathbf{1})\rangle\langle \ell(v_1), v_j\rangle
\end{aligned}$$

If we define a measure μ on Γ by $\mu_i = \rho(\mathbf{1})_i \ell(v_1)_i$ (where subscripts denote vector components) then it follows that the map $X \to \Gamma$ taking each walk to its initial vertex pushes forward the measure μ on X to the measure μ on Γ.

Define a matrix N with entries

$$N_{ij} = \frac{M_{ij}\rho(\mathbf{1})_j}{\lambda\rho(\mathbf{1})_i}$$

if $\rho(\mathbf{1})_i$ is nonzero, and set $N_{ii} = 1$ and $N_{ij} = 0$ otherwise. Recall that a non-negative matrix N with the property that $\sum_j N_{ij} = 1$ for any i is called a *stochastic matrix* (compare with the matrix N in § 6.1.4).

LEMMA 6.66. *The matrix N is stochastic, and satisfies $\mu N = \mu$.*

PROOF. For any i not in the support of $\rho(\mathbf{1})$, we have $\sum_j N_{ij} = 1$ by fiat. Otherwise,

$$\sum_j N_{ij} = \sum_j \frac{M_{ij}\rho(\mathbf{1})_j}{\lambda\rho(\mathbf{1})_i} = \frac{(M\rho(\mathbf{1}))_i}{\lambda\rho(\mathbf{1})_i} = 1$$

This shows N is a stochastic matrix. To verify the second formula,

$$\begin{aligned}
\sum_i \mu_i N_{ij} &= \sum_i \rho(\mathbf{1})_i \ell(v_1)_i \frac{M_{ij}\rho(\mathbf{1})_j}{\lambda\rho(\mathbf{1})_i} \\
&= \frac{1}{\lambda}\rho(\mathbf{1})_j \sum_i \ell(v_1)_i M_{ij} \\
&= \rho(\mathbf{1})_j \ell(v_1)_j = \mu_j
\end{aligned}$$

where the sum is over i with $\mu_i \neq 0$ which implies $\rho(\mathbf{1})_i \neq 0$. ∎

By a further abuse of notation, we let $p : X \to X_n$ denote the restriction of an infinite path to a suitable prefix. We can obtain a formula for the measure μ on cylinders $p^{-1}(x) \subset X$ for $x \in X_n$ in terms of the measure μ on Γ, and the matrix N.

LEMMA 6.67. *For $x \in X_n$, there is equality*

$$\mu(p^{-1}(x)) = \mu_{i_0} N_{i_0 i_1} N_{i_1 i_2} \cdots N_{i_{n-1} i_n}$$

where $x = (x_{i_0}, x_{i_1}, \cdots, x_{i_n})$, and x_{i_j} corresponds to the vertex v_{i_j} of Γ.

PROOF. Let $g \in G_n$. If γ_g is the corresponding walk in Γ, let v_i be the last vertex of γ_g. Then $\widehat{\nu}(g) = \lambda^{-n}\rho(\mathbf{1})_i$. Moreover, for each vertex v_j, there are M_{ij} elements $h \in \text{cone}(g)$ for which the corresponding walks γ_h have last vertex v_j. Each h has $\widehat{\nu}(h) = \lambda^{-n-1}\rho(\mathbf{1})_j$ so given g, the sum over $h \in G_{n+1}$ with $h \in \text{cone}(g)$ which have last vertex v_j of $\widehat{\nu}(h)$ is $M_{ij}\rho(\mathbf{1})_j/\lambda\rho(\mathbf{1})_i = N_{ij}$. In other words, given any $y \in Y$ whose nth vertex is v_i, the probability in the $\widehat{\nu}$ measure that its $(n+1)$st vertex is v_j is N_{ij}. Since this formula does not depend on n but just v_i and v_j, the lemma is proved. $\square$

We call μ on Γ the *stationary measure*. It is not necessarily a probability measure, but it determines a unique probability measure by scaling. By abuse of notation, we refer to these two measures by the same name. Lemma 6.67 may be interpreted as saying that a random walk on Γ with initial vertex chosen randomly with respect to the stationary measure μ and with transition probabilities given by the stochastic matrix N agrees with a random element of X with respect to the measure μ.

The next Lemma describes the support of the stationary measure μ on Γ.

LEMMA 6.68. *The support of the stationary measure is equal to the disjoint union of the maximal recurrent subgraphs C^i of Γ whose adjacency matrices have biggest eigenvalue λ.*

PROOF. Since $\mu_i = \rho(\mathbf{1})_i \ell(v_1)_i$ a vertex v_i is in the support of μ_i if for some large fixed k there are $\Theta(\lambda^n)$ paths of length between n and $n+k$ from v_1 to v_i, and $\Theta(\lambda^n)$ paths of length n from v_i to some other vertex. It follows that some path from v_1 to v_i intersects a maximal recurrent component C whose adjacency matrix has biggest real eigenvalue $\xi(C) = \lambda$, and similarly there is some outgoing path from v_i which intersects a maximal recurrent component C' with $\xi(C') = \lambda$. Lemma 6.58 implies that $C = C'$, and therefore $v_i \in C$.

Conversely, let C be a recurrent subgraph of Γ whose adjacency matrix has eigenvalue λ. Then $\rho(\mathbf{1})_i$ and $\ell(v_1)_i$ are positive for all v_i in C, by counting only paths which stay in C outside a prefix and suffix of bounded length. $\square$

From the point of view of stationary measure, Γ decomposes into a finite union of recurrent subgraphs C^i, each with Perron–Frobenius eigenvalue λ. Let $N|_{C^i}$ denote the restriction of the stochastic matrix N to the subgraph C^i. Then $N|_{C^i}$ is a stochastic matrix. Let μ^i denote the measure μ on Γ restricted to C^i, and rescaled to be a probability measure. Then $N|_{C^i}$ preserves μ^i, and determines an *ergodic stationary Markov chain* on the vertices of C^i.

Let ϕ be weakly combable. As in § 6.1.4 we can define $\overline{S}_n^i$ to be equal to the sum of the values of $d\phi$ on a random walk on C^i of length n with respect to the stationary measure μ^i and transition probabilities given by $N|_{C^i}$.

The Central Limit Theorem for ergodic stationary Markov chains implies

LEMMA 6.69. *Let ϕ be weakly combable. With terminology as above, there is convergence in the sense of distribution*

$$\lim_{n\to\infty} n^{-1/2}(\overline{S}_n^i - nE^i) \to N(0,\sigma^i)$$

for some $\sigma^i \geq 0$ where E^i denotes the average of $d\phi$ on C^i with respect to the stationary measure μ^i, and $N(0,\sigma^i)$ denotes the Gaussian normal distribution with mean 0 and standard deviation σ^i.

This theorem is essentially due to Markov [**146**]. For a proof and more details, as well as a precise formula for σ, see e.g. [**179**], Chapter 4, § 46 or [**96**], § 11.5, especially Theorem 11.17. An excellent general reference is [**127**].

REMARK 6.70. Note that $\sigma = 0$ is possible (for instance, ϕ could be identically zero), in which case by convention, $N(0,\sigma)$ denotes the Dirac distribution with mass 1 centered at 0.

6.5.4. Central Limit Theorem. In order to derive a central limit theorem for the group G as a whole, we must compare the means E^i and standard deviations σ^i associated to distinct components C^i.

For each component C^i in the support of the stationary measure μ, let $Y^i \subset Y$ denote the set of infinite paths in Γ which eventually enter C^i and stay there. Note that the Y^i are disjoint, and $\widehat{\nu}(Y - \cup_i Y^i) = 0$. For each path $\gamma \in Y$ we can consider the following. Let $\gamma_i \in G$ be the element corresponding to the evaluation of the word which is equal to the prefix of γ of length i. We fix the following notation: if r is a real number, let $\delta(r)$ denote the probability measure on $\mathbb{R}$ which consists of an atom concentrated at r. For a given real number A, and for integers n, m we can consider the following measure.

$$\omega(n,m)(\gamma) = \sum_{i=1}^{m} \frac{1}{m}\,\delta\left((\phi(\gamma_{i+n}) - \phi(\gamma_i) - nA)n^{-1/2}\right)$$

and then define $\omega(\gamma) = \lim_{n\to\infty}\lim_{m\to\infty}\omega(n,m)(\gamma)$. Note that the existence of this limit depends on the "correct" choice of A.

DEFINITION 6.71. Let $\gamma \in Y^i$. We say that γ is *typical* if $\omega(\gamma)$ exists for $A = E^i$, and is equal to $N(0,\sigma^i)$. More generally, if γ is an infinite geodesic ray in G, then γ is E,σ-typical if $\omega(\gamma)$ exists for $A = E$ and is equal to $N(0,\sigma)$.

From Lemma 6.69 we obtain the following, which does not depend on ϕ being bicombable, but only weakly combable:

LEMMA 6.72. *Almost every $\gamma \in Y^i$ with respect to the measure $\widehat{\nu}$ is E^i,σ^i-typical.*

PROOF. The following proof was suggested by Shigenori Matsumoto.

We fix the notation below for the course of the Lemma (the reader should be warned that it is slightly incompatible with notation used elsewhere; this is done to avoid a proliferation of subscripts). Let C_i be a component of Γ with Perron–Frobenius root $\xi(C_i) = \lambda$. Let Y_i be the set of infinite paths in Γ starting at v_1 that eventually stay in C_i, and let X_i be the set of infinite paths in C_i. There is a measure $\widehat{\mu}_i$ on X_i obtained by restricting μ on X. The measure $\widehat{\mu}_i$ is determined by a stationary measure μ_i on C_i and the transition matrix $N(i)$, the restriction of the measures μ and the matrix N defined in § 6.5.3. The measure μ_i is stationary in the sense that $\mu_i^T N(i) = \mu_i^T$, so $\widehat{\mu}_i$ is shift invariant. Since C_i is recurrent, μ_i^T is the only eigenvector of $N(i)$ with eigenvalue 1, so μ_i is extremal in the space of stationary measures. Therefore by the random ergodic theorem (see e.g. [**168**], Ch. 10) the measure $\widehat{\mu}_i$ on X_i is ergodic.

Now, there is a subset X_i^* of X_i of full measure such that for all $\gamma \in X_i^*$,

$$\frac{1}{m}\sum_0^m \delta_{S^k\gamma} \to \widehat{\mu}_i$$

in the weak* topology, where S denotes shift map, and δ is a Dirac mass. On the other hand, on Y_i there is a measure $\widehat{\nu}_i$ which is the restriction of $\widehat{\nu}$. Define $q : Y_i \to X_i$ by

$$q(\gamma) = S^{n(\gamma)}(\gamma)$$

where $n : Y_i \to \mathbb{N}$ satisfies the following condition. Let $\pi : X_i \to C_i$ take each infinite walk to its initial vertex. Choose n so that $\pi \circ q : Y_i \to C_i$ sends the measure $\widehat{\nu}_i$ on Y_i to a measure μ_q on C_i of full support. The measure $q_*\widehat{\nu}_i$ on X_i is obtained from an initial measure μ_q and the transition matrix $N(i)$ as in § 6.5.3; it follows that the measures $q_*\widehat{\nu}_i$ and μ_i are equivalent (i.e. each is absolutely continuous with respect to the other).

It follows that $Y_i^* := q^{-1}(X_i^*)$ has full measure with respect to $\widehat{\nu}_i$, and if $\gamma \in Y_i^*$, then

$$\frac{1}{m}\sum_0^m \delta_{S^k\gamma} \to \widehat{\mu}_i$$

This shows that the geodesic ray in G associated to any $\gamma \in Y_i^*$ is E^i, σ^i-typical, and the lemma is proved. □

On the other hand, the following Lemma uses bicombability in an essential way:

LEMMA 6.73. *Let γ be an E, σ-typical geodesic ray in G. If ϕ is combable and if γ' is a geodesic ray with the same endpoint at γ, then γ' is also E, σ-typical. If ϕ is bicombable then for any $g \in G$, the translate $g\gamma$ is E, σ-typical.*

PROOF. Let γ and γ' have the same endpoint. Then there is a constant C such that $d_L(\gamma_i, \gamma_i') \le C$ and therefore $|\phi(\gamma_i) - \phi(\gamma_i')| \le K$ for some K independent of i. This shows that γ' is E, σ-typical if γ is. Similarly, if $g \in G$ then $d_R(g\gamma_i, \gamma_i) \le C$ and therefore $|\phi(g\gamma_i) - \phi(\gamma_i)| \le K$ for some K independent of i. □

We now come to the crucial point. For each i, let $\partial^i G$ denote the image of the typical elements in Y^i under the endpoint map $Y \to \partial G$. Note that $\widehat{\nu}(\partial^i G)$ is strictly positive for each i. By Theorem 6.63, for any i, j there is some $g \in G$ with $\widehat{\nu}(g\partial^i G \cap \partial^j G) > 0$. It follows that there is a typical $\gamma \in Y^i$ and a typical $\gamma' \in Y^j$ such that (identifying elements of Y and geodesic rays in G starting at id) the translate $g\gamma$ and γ' are asymptotic to the same endpoint in ∂G. Since ϕ is bicombable, by Lemma 6.73, γ and γ' are both E^i, σ^i-typical and E^j, σ^j-typical. It follows that $E^i = E^j$ and $\sigma^i = \sigma^j$. Together with Lemma 6.69, this proves the Central Limit Theorem:

THEOREM 6.74 (Central Limit Theorem; Calegari–Fujiwara, [**50**]). *Let ϕ be a bicombable function on a word-hyperbolic group G. Let $\overline{\phi}_n$ be the value of ϕ on a random word of length n with respect to the $\widehat{\nu}$ measure. Then there is convergence in the sense of distribution*

$$\lim_{n\to\infty} n^{-1/2}(\overline{\phi}_n - nE) \to N(0,\sigma)$$

for some $\sigma \ge 0$, where E denotes the average of $d\phi$ on Γ with respect to the stationary measure.

The following corollary does not make reference to the measure $\widehat{\nu}$.

COROLLARY 6.75. *Let ϕ be a bicombable function on a word-hyperbolic group G. Then there is a constant E such that for any $\epsilon > 0$ there is a K and an N so that if G_n denotes the set of elements of length $n \ge N$, there is a subset G'_n with $|G'_n|/|G_n| \ge 1-\epsilon$, so that for all $g \in G'_n$, there is an inequality*

$$|\phi(g) - nE| \le K \cdot \sqrt{n}$$

As a special case, let S_1, S_2 be two finite symmetric generating sets for G. Word length in the S_2 metric is a bicombable function with respect to a combing L_1 for the S_1 generating set. Hence:

COROLLARY 6.76. *Let S_1 and S_2 be finite generating sets for G. There is a constant $\lambda_{1,2}$ such that for any $\epsilon > 0$, there is a K and an N so that if G_n denotes the set of elements of length $n \ge N$ in the S_1 metric, there is a subset G'_n with $|G'_n|/|G_n| \ge 1-\epsilon$, so that for all $g \in G'_n$ there is an equality*

$$\big|\lambda_{1,2}|g|_{S_1} - |g|_{S_2}\big| = \big|\lambda_{1,2}\cdot n - |g|_{S_2}\big| \le K\cdot\sqrt{n}$$

REMARK 6.77. In Corollary 6.76 it is important to note that though a typical geodesic word of length n in the S_1 metric is represented by a geodesic word of length $n\cdot\lambda_{1,2}$ in the S_2 metric, with error of order $\sqrt{n}$, the resulting set of geodesic words in the S_2 metric are not themselves typical. Thus $\lambda_{1,2}\lambda_{2,1} > 1$ in general. We give an example to illustrate this phenomenon in § 6.5.5.

If ϕ is a quasimorphism, then $|\phi(g) + \phi(g^{-1})| \le$ const. so if S is symmetric, then necessarily E as above is equal to 0. Hence:

COROLLARY 6.78. *Let ϕ be a bicombable quasimorphism on a word-hyperbolic group G. Let $\overline{\phi}_n$ be the value of ϕ on a random word of length n with respect to the $\widehat{\nu}$ measure. Then there is convergence in the sense of distribution*

$$\lim_{n\to\infty} n^{-1/2}\overline{\phi}_n \to N(0,\sigma)$$

for some $\sigma \ge 0$.

6.5.5. An example. Let F denote the free group on two generators a, b. Let S_1 denote the symmetric generating set $S_1 = \langle a, b, a^{-1}, b^{-1}\rangle$ and S_2 the symmetric generating set $S_2 = \langle a, b, c, a^{-1}, b^{-1}, c^{-1}\rangle$ where $c = ab$ (and therefore $c^{-1} = b^{-1}a^{-1}$). We compare word length in the S_1 and the S_2 metrics.

One can verify that a word in the S_2 generating set is a geodesic if and only if it is reduced, and contains no subwords of the form $a^{-1}c, cb^{-1}, c^{-1}a, c^{-1}b$, and moreover that geodesic representatives are unique. The language of all geodesics in the S_2 generating set is therefore a combing.

One can build a digraph Γ which parameterizes the language of geodesics in S_2 as follows. There are seven vertices, one initial vertex and six other vertices labeled by the elements of S_2. There is an outgoing edge from the initial vertex to each other vertex, and one directed edge from x to y for each other vertex if and only if xy is not one of the four "excluded" words above. See Figure 6.7. The vertices have been labeled a, b, c, A, B, C and labels have been left off the edges for clarity.

Let Γ' be obtained from Γ by removing the initial vertex. The adjacency matrix of Γ' is

$$M = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

which is Perron–Frobenius with biggest real eigenvalue 4, and $\mathbf{1}^T, \mathbf{1}$ as left and right eigenvectors. It follows that the stationary measure is just equal to the ordinary uniform measure. Note that there are $6 \times 4^{n-1}$ words of length n in the S_2 metric, and $4 \times 3^{n-1}$ words of length n in the S_1 metric.

Let ϕ_{S_i} denote the bicombable function which computes word length in the S_i metric. There are discrete derivatives $d\phi_{S_1}, d\phi_{S_2}$ from the vertices of Γ' to 1. Here $d\phi_{S_2}$ is just the constant function $\Gamma' \to 1$, whereas $d\phi_{S_1}$ takes the value 1 on the vertices labeled a, b, A, B and 2 on the vertices labeled c, C. It follows that a random word of length n in the S_2 metric has length $4n/3$ in the S_1 metric, with error of order $\sqrt{n}$.

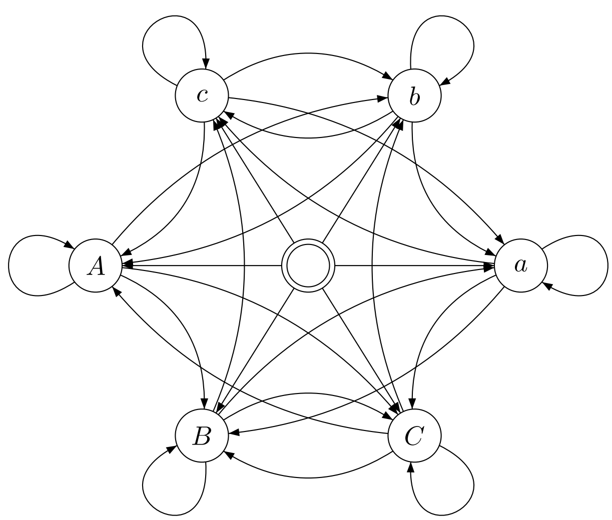


FIGURE 6.7. A digraph parameterizing geodesics in the S_2 metric

On the other hand, $d\phi_{S_1}$ and $d\phi_{S_2}$ exist as functions from the vertices of Γ_1' to 1 where Γ_1' is the digraph in Figure 6.4. In this case, $d\phi_{S_1}$ is the constant function $\Gamma_1' \to 1$ and $d\phi_{S_2}$ is the function which takes the value 0 on the vertices labeled ab and $b^{-1}a^{-1}$, and 1 on all other vertices. It follows that a random word of length n in the S_1 metric has length $5n/6$ in the S_2 metric, with error of order $\sqrt{n}$. Hence $\lambda_{1,2}\lambda_{2,1} = 5/6 \times 4/3 = 10/9$. with notation as in Corollary 6.76.

In general, if the growth rate in the S_i metric is λ_i^n for $i = 1, 2$ then there is an inequality $\lambda_{i,j} \geq \log \lambda_i / \log \lambda_j$, by counting. In this case, we get the two (easily verified) inequalities

$$0.83333 \cdots = \frac{5}{6} \geq \frac{\log 3}{\log 4} = 0.79248 \cdots$$

and

$$1.33333 \cdots = \frac{4}{3} \geq \frac{\log 4}{\log 3} = 1.26186 \cdots$$

REMARK 6.79. The numbers $\lambda_{1,2}$ where S_1 and S_2 are a symmetric basis for a free group F_k, are studied in [**119**], where it is shown that they are always rational, and satisfy $2k\lambda \in \mathbb{Z}[1/(2k-1)]$.

Bibliography

[1] Vladimir I. Arnol′d, *Certain topological invariants of algebraic functions*, Trudy Moskov. Mat. Obšč. **21** (1970), 27–46.

[2] ———, *Geometrical methods in the theory of ordinary differential equations*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science], vol. 250, Springer-Verlag, New York, 1983, Translated from the Russian by Joseph Szücs, Translation edited by Mark Levi.

[3] ———, *Sturm theorems and symplectic geometry*, Funktsional. Anal. i Prilozhen. **19** (1985), no. 4, 1–10, 95.

[4] ———, *The asymptotic Hopf invariant and its applications*, Selecta Math. Soviet. **5** (1986), no. 4, 327–345, Selected translations.

[5] Emil Artin, *Theorie der Zöpfe*, Abh. Math. Sem. Hamburgischen Univ. **4** (1926), 47–72.

[6] Jean Barge and Étienne Ghys, *Surfaces et cohomologie bornée*, Invent. Math. **92** (1988), no. 3, 509–526.

[7] ———, *Cocycles d'Euler et de Maslov*, Math. Ann. **294** (1992), no. 2, 235–265.

[8] Christophe Bavard, *Longueur stable des commutateurs*, Enseign. Math. (2) **37** (1991), no. 1-2, 109–150.

[9] Gennadiĭ V. Belyĭ, *Galois extensions of a maximal cyclotomic field*, Izv. Akad. Nauk SSSR Ser. Mat. **43** (1979), no. 2, 267–276, 479.

[10] Riccardo Benedetti and Carlo Petronio, *Lectures on hyperbolic geometry*, Universitext, Springer-Verlag, Berlin, 1992.

[11] Abraham Berman and Robert J. Plemmons, *Nonnegative matrices in the mathematical sciences*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1979, Computer Science and Applied Mathematics.

[12] Mladen Bestvina and Mark Feighn, *A hyperbolic* $\mathrm{Out}(F_n)$ *complex*, 2008, arXiv:0808.3730.

[13] Mladen Bestvina and Koji Fujiwara, *Bounded cohomology of subgroups of mapping class groups*, Geom. Topol. **6** (2002), 69–89 (electronic).

[14] Robert Bieri and Ralph Strebel, *On groups of* PL *homeomorphisms of the real line*, preprint, 1985.

[15] Stephen Bigelow, *The Burau representation is not faithful for* $n = 5$, Geom. Topol. **3** (1999), 397–404 (electronic).

[16] Joan S. Birman, *Braids, links, and mapping class groups*, Princeton University Press, Princeton, N.J., 1974, Annals of Mathematics Studies, No. 82.

[17] Béla Bollobás and Oliver Riordan, *A polynomial invariant of graphs on orientable surfaces*, Proc. London Math. Soc. (3) **83** (2001), no. 3, 513–531.

[18] Michael D. Boshernitzan, *Dense orbits of rationals*, Proc. Amer. Math. Soc. **117** (1993), no. 4, 1201–1203.

[19] Abdessalam Bouarich, *Suites exactes en cohomologie bornée réelle des groupes discrets*, C. R. Acad. Sci. Paris Sér. I Math. **320** (1995), no. 11, 1355–1359.

[20] Brian H. Bowditch, *A topological characterisation of hyperbolic groups*, J. Amer. Math. Soc. **11** (1998), no. 3, 643–667.

[21] ———, *Tight geodesics in the curve complex*, Invent. Math. **171** (2008), no. 2, 281–300.

[22] Steven Boyer, Dale Rolfsen, and Bert Wiest, *Orderable 3-manifold groups*, Ann. Inst. Fourier (Grenoble) **55** (2005), no. 1, 243–288.

[23] Volker Braungardt and Dieter Kotschick, *Clustering of critical points in Lefschetz fibrations and the symplectic Szpiro inequality*, Trans. Amer. Math. Soc. **355** (2003), no. 8, 3217–3226 (electronic).

[24] Martin R. Bridson and André Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 319, Springer-Verlag, Berlin, 1999.

[25] Nathan Broaddus, Benson Farb, and Andrew Putman, *The Casson invariant and the word metric on the Torelli group*, C. R. Math. Acad. Sci. Paris **345** (2007), no. 8, 449–452.

[26] Sergei D. Brodskiĭ, *Equations over groups, and groups with one defining relation*, Sibirsk. Mat. Zh. **25** (1984), no. 2, 84–103.

[27] Robert Brooks, *Some remarks on bounded cohomology*, Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978) (Princeton, N.J.), Ann. of Math. Stud., vol. 97, Princeton Univ. Press, 1981, pp. 53–63.

[28] Kenneth S. Brown, *Finiteness properties of groups*, Proceedings of the Northwestern conference on cohomology of groups (Evanston, Ill., 1985), vol. 44, 1987, pp. 45–75.

[29] Dmitri Burago, Sergei Ivanov, and Leonid Polterovich, *Conjugation-invariant norms on groups of geometric origin*, Groups of Diffeomorphisms: In Honor of Shigeyuki Morita on the Occasion of His 60th Birthday, Adv. Stud. Pure Math., vol. 52, Kinokuniya, Tokyo, 2008, pp. 221–250.

[30] Marc Burger and Alessandra Iozzi, *Bounded Kähler class rigidity of actions on Hermitian symmetric spaces*, Ann. Sci. École Norm. Sup. (4) **37** (2004), no. 1, 77–103.

[31] Marc Burger, Alessandra Iozzi, and Anna Wienhard, *Surface group representations with maximal Toledo invariant*, C. R. Math. Acad. Sci. Paris **336** (2003), no. 5, 387–390.

[32] ———, *Surface group representations with maximal Toledo invariant*, preprint, to appear in Annals of Mathematics, 2008.

[33] Marc Burger and Nicolas Monod, *Bounded cohomology of lattices in higher rank Lie groups*, J. Eur. Math. Soc. (JEMS) **1** (1999), no. 2, 199–235.

[34] ———, *Continuous bounded cohomology and applications to rigidity theory*, Geom. Funct. Anal. **12** (2002), no. 2, 219–280.

[35] ———, *On and around the bounded cohomology of* SL_2, Rigidity in dynamics and geometry (Cambridge, 2000), Springer, Berlin, 2002, pp. 19–37.

[36] Robert G. Burns and Victor W. D. Hale, *A note on group rings of certain torsion-free groups*, Canad. Math. Bull. **15** (1972), 441–445.

[37] Peter Buser, *Geometry and spectra of compact Riemann surfaces*, Progress in Mathematics, vol. 106, Birkhäuser Boston Inc., Boston, MA, 1992.

[38] Eugenio Calabi, *On the group of automorphisms of a symplectic manifold*, Problems in analysis (Lectures at the Sympos. in honor of Salomon Bochner, Princeton Univ., Princeton, N.J., 1969), Princeton Univ. Press, Princeton, N.J., 1970, pp. 1–26.

[39] Danny Calegari, `scallop`, computer program; available from `http://www.its.caltech.edu/~dannyc/scl/scallop.tar.gz`.

[40] ———, *Foliations and the geometry of* 3*-manifolds*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 2007, Oxford Science Publications.

[41] ———, *Stable commutator length in subgroups of* $\mathrm{PL}^+(I)$, Pacific J. Math. **232** (2007), no. 3, 257–262.

[42] ———, *Length and stable length*, Geom. Funct. Anal. **18** (2008), no. 1, 50–76.

[43] ———, *Surface subgroups from homology*, Geom. Topol. **12** (2008), 1995–2007 (electronic).

[44] ———, *Word length in surface groups with characteristic generating sets*, Proc. Amer. Math. Soc. **136** (2008), no. 7, 2631–2637.

[45] ———, *Faces of the scl norm ball*, Geom. Topol. **13** (2009), 1313–1336 (electronic).

[46] ———, *Sails, surgery and scl*, preprint, 2009.

[47] ———, *Stable commutator length is rational in free groups*, Jour. Amer. Math. Soc., to appear, 2009.

[48] Danny Calegari and Nathan M. Dunfield, *Laminations and groups of homeomorphisms of the circle*, Invent. Math. **152** (2003), no. 1, 149–204.

[49] Danny Calegari and Koji Fujiwara, *Stable commutator length in word hyperbolic groups*, Groups, Geom., Dynam., to appear.

[50] ———, *Combable functions, quasimorphisms, and the Central Limit Theorem*, arXiv: 0805.1755, 2008.

[51] James W. Cannon, *The combinatorial structure of cocompact discrete hyperbolic groups*, Geom. Dedicata **16** (1984), no. 2, 123–148.

[52] James W. Cannon, William J. Floyd, and Walter R. Parry, *Introductory notes on Richard Thompson's groups*, Enseign. Math. (2) **42** (1996), no. 3-4, 215–256.
[53] David Carter and Gordon Keller, *Bounded elementary generation of* $\mathrm{SL}_n(\mathcal{O})$, Amer. J. Math. **105** (1983), no. 3, 673–687.
[54] Jean-Louis Clerc and Khalid Koufany, *Primitive du cocycle de Maslov généralisé*, Math. Ann. **337** (2007), no. 1, 91–138.
[55] Jean-Louis Clerc and Bent Ørsted, *The Gromov norm of the Kaehler class and the Maslov index*, Asian J. Math. **7** (2003), no. 2, 269–295.
[56] Michel Coornaert, *Mesures de Patterson-Sullivan sur le bord d'un espace hyperbolique au sens de Gromov*, Pacific J. Math **159** (1993), no. 2, 241–270.
[57] Michel Coornaert and Athanase Papadopoulos, *Symbolic dynamics and hyperbolic groups*, Lecture Notes in Mathematics, vol. 1539, Springer-Verlag, Berlin, 1993.
[58] David Coppola, Harriett Purves, Allison McCoy, and Dale Purves, *The distribution of oriented contours in the real world*, Proc. Natl. Acad. Sci. USA **95** (1998), no. 7, 4002–4006.
[59] Marc Culler, *Using surfaces to solve equations in free groups*, Topology **20** (1981), no. 2, 133–145.
[60] Marc Culler and Peter B. Shalen, *Varieties of group representations and splittings of* 3-*manifolds*, Ann. of Math. (2) **117** (1983), no. 1, 109–146.
[61] Marc Culler and Karen Vogtmann, *Moduli of graphs and automorphisms of free groups*, Invent. Math. **84** (1986), no. 1, 91–119.
[62] George B. Dantzig, *Linear programming and extensions*, Princeton University Press, Princeton, N.J., 1963.
[63] Max Dehn, *Papers on group theory and topology*, Springer-Verlag, New York, 1987, Translated from the German and with introductions and an appendix by John Stillwell, With an appendix by Otto Schreier.
[64] Thomas Delzant, *Sous-groupes à deux générateurs des groupes hyperboliques*, Group theory from a geometrical viewpoint (Trieste, 1990), World Sci. Publ., River Edge, NJ, 1991, pp. 177–189.
[65] ———, *Sous-groupes distingués et quotients des groupes hyperboliques*, Duke Math. J. **83** (1996), no. 3, 661–682.
[66] Antun Domic and Domingo Toledo, *The Gromov norm of the Kaehler class of symmetric domains*, Math. Ann. **276** (1987), no. 3, 425–432.
[67] Andrew J. Duncan and James Howie, *The genus problem for one-relator products of locally indicable groups*, Math. Z. **208** (1991), no. 2, 225–237.
[68] Charles C. Edmunds, *On the endomorphism problem for free groups*, Comm. Algebra **3** (1975), 1–20.
[69] ———, *On the endomorphism problem for free groups. II*, Proc. London Math. Soc. (3) **38** (1979), no. 1, 153–168.
[70] David Eisenbud, Ulrich Hirsch, and Walter Neumann, *Transverse foliations of Seifert bundles and self-homeomorphism of the circle*, Comment. Math. Helv. **56** (1981), no. 4, 638–660.
[71] Yakov Eliashberg and Leonid Polterovich, *Partially ordered groups and geometry of contact transformations*, Geom. Funct. Anal. **10** (2000), no. 6, 1448–1476.
[72] Erich W. Ellers and Nikolai Gordeev, *On the conjectures of J. Thompson and O. Ore*, Trans. Amer. Math. Soc. **350** (1998), no. 9, 3657–3671.
[73] Hisaaki Endo and Dieter Kotschick, *Bounded cohomology and non-uniform perfection of mapping class groups*, Invent. Math. **144** (2001), no. 1, 169–175.
[74] ———, *Failure of separation by quasi-homomorphisms in mapping class groups*, Proc. Amer. Math. Soc. **135** (2007), no. 9, 2747–2750 (electronic).
[75] Michael Entov and Leonid Polterovich, *Calabi quasimorphism and quantum homology*, Int. Math. Res. Not. (2003), no. 30, 1635–1676.
[76] Michael Entov, Leonid Polterovich, and Pierre Py, *On continuity of quasi-morphisms for symplectic maps*, arXiv: 0904.1397, 2009.
[77] David B. A. Epstein, James W. Cannon, Derek F. Holt, Silvio V. F. Levy, Michael S. Paterson, and William P. Thurston, *Word processing in groups*, Jones and Bartlett Publishers, Boston, MA, 1992.
[78] David B. A. Epstein and Koji Fujiwara, *The second bounded cohomology of word-hyperbolic groups*, Topology **36** (1997), no. 6, 1275–1289.

[79] Benson Farb, Christopher J. Leininger, and Dan Margalit, *The lower central series and pseudo-Anosov dilatations*, Amer. J. Math. **130** (2008), no. 3, 799–827.

[80] Benson Farb and Dan Margalit, *A primer on mapping class groups*, preprint; available from the authors' websites, 2007.

[81] Gordon M. Fisher, *On the group of all homeomorphisms of a manifold*, Trans. Amer. Math. Soc. **97** (1960), 193–212.

[82] Koji Fujiwara, *The second bounded cohomology of a group acting on a Gromov-hyperbolic space*, Proc. London Math. Soc. (3) **76** (1998), no. 1, 70–94.

[83] ———, *The second bounded cohomology of an amalgamated free product of groups*, Trans. Amer. Math. Soc. **352** (2000), no. 3, 1113–1129.

[84] David Gabai, *Foliations and the topology of* 3*-manifolds*, J. Differential Geom. **18** (1983), no. 3, 445–503.

[85] Jean-Marc Gambaudo and Étienne Ghys, *Enlacements asymptotiques*, Topology **36** (1997), no. 6, 1355–1379.

[86] ———, *Commutators and diffeomorphisms of surfaces*, Ergodic Theory Dynam. Systems **24** (2004), no. 5, 1591–1617.

[87] ———, *Braids and signatures*, Bull. Soc. Math. France **133** (2005), no. 4, 541–579.

[88] Jean-Marc Gambaudo, Dennis Sullivan, and Charles Tresser, *Infinite cascades of braids and smooth dynamical systems*, Topology **33** (1994), no. 1, 85–94.

[89] Alexander Gel'fond, *On Hilbert's seventh problem*, Dokl. Akad. Nauk SSSR (N.S.) **2** (1934), no. 1-6, 177–182.

[90] Steven M. Gersten, *Cohomological lower bounds for isoperimetric functions on groups*, Topology **37** (1998), no. 5, 1031–1072.

[91] Étienne Ghys, *Knots and dynamics*, International Congress of Mathematicians. Vol. I, Eur. Math. Soc., Zürich, 2007, pp. 247–277.

[92] Étienne Ghys and Vlad Sergiescu, *Sur un groupe remarquable de difféomorphismes du cercle*, Comment. Math. Helv. **62** (1987), no. 2, 185–239.

[93] William M. Goldman, *Discontinuous groups and the Euler class*, Ph.D. thesis, UC Berkeley, 1980.

[94] Cameron Gordon and Henry Wilton, *On surface subgroups of doubles of free groups*, arXiv: 0902.3693, 2009.

[95] Rotislav I. Grigorchuk, *Some results on bounded cohomology*, Combinatorial and geometric group theory (Edinburgh, 1993), London Math. Soc. Lecture Note Ser., vol. 204, Cambridge Univ. Press, Cambridge, 1995, pp. 111–163.

[96] Charles M. Grinstead and J. Laurie Snell, *Introduction to Probability*, second ed., American Mathematical Society, Providence, RI, 1997.

[97] Mikhael Gromov, *Volume and bounded cohomology*, Inst. Hautes Études Sci. Publ. Math. (1982), no. 56, 5–99 (1983).

[98] ———, *Hyperbolic groups*, Essays in group theory, Math. Sci. Res. Inst. Publ., vol. 8, Springer, New York, 1987, pp. 75–263.

[99] ———, *Asymptotic invariants of infinite groups*, Geometric group theory, Vol. 2 (Sussex, 1991), London Math. Soc. Lecture Note Ser., vol. 182, Cambridge Univ. Press, Cambridge, 1993, pp. 1–295.

[100] Alexandre Grothendieck, *Esquisse d'un programme*, Geometric Galois actions, 1, London Math. Soc. Lecture Note Ser., vol. 242, Cambridge Univ. Press, Cambridge, 1997, With an English translation on pp. 243–283, pp. 5–48.

[101] Uffe Haagerup and Hans J. Munkholm, *Simplices of maximal volume in hyperbolic n-space*, Acta Math. **147** (1981), no. 1-2, 1–11.

[102] Marshall Hall, Jr., *Coset representations in free groups*, Trans. Amer. Math. Soc. **67** (1949), 421–432.

[103] William J. Harvey, *Boundary structure of the modular group*, Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978) (Princeton, N.J.), Ann. of Math. Stud., vol. 97, Princeton Univ. Press, 1981, pp. 245–251.

[104] Joel Hass and Peter Scott, *Configurations of curves and geodesics on surfaces*, Proceedings of the Kirbyfest (Berkeley, CA, 1998), Geom. Topol. Monogr., vol. 2, Geom. Topol. Publ., Coventry, 1999, pp. 201–213 (electronic).

[105] Allen Hatcher, *On triangulations of surfaces*, Topology Appl. **40** (1991), no. 2, 189–194.

[106] Allen Hatcher and William P. Thurston, *A presentation for the mapping class group of a closed orientable surface*, Topology **19** (1980), no. 3, 221–237.

[107] John Hempel, 3-*Manifolds*, Princeton University Press, Princeton, N. J., 1976, Ann. of Math. Studies, No. 86.

[108] ———, *3-manifolds as viewed from the curve complex*, Topology **40** (2001), no. 3, 631–657.

[109] Morris W. Hirsch and William P. Thurston, *Foliated bundles, invariant measures and flat manifolds*, Ann. Math. (2) **101** (1975), 369–390.

[110] Gerhard Hochschild and Jean-Pierre Serre, *Cohomology of group extensions*, Trans. Amer. Math. Soc. **74** (1953), 110–134.

[111] Craig D. Hodgson and Steven P. Kerckhoff, *Universal bounds for hyperbolic Dehn surgery*, Ann. of Math. (2) **162** (2005), no. 1, 367–421.

[112] Helmut Hofer and Eduard Zehnder, *Symplectic invariants and Hamiltonian dynamics*, Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks], Birkhäuser Verlag, Basel, 1994.

[113] Matthew Horsham and Richard Sharp, *Lengths, quasi-morphisms and statistics for free groups*, Spectral Analysis in Geometry and Number Theory, Contemp. Math., vol. 484, American Mathematical Society, Providence, RI, 2009, pp. 219–238.

[114] James Howie, *On locally indicable groups*, Math. Zeit. **180** (1982), no. 4, 445–461.

[115] Nikolai V. Ivanov, *Foundations of the theory of bounded cohomology*, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **143** (1985), 69–109, 177–178, Studies in topology, V.

[116] Mogens Høgh Jensen, Per Bak, and Tomas Bohr, *Transition to chaos by interaction of resonances in dissipative systems. I. Circle maps*, Phys. Rev. A (3) **30** (1984), no. 4, 1960–1969.

[117] Barry Edward Johnson, *Cohomology in Banach algebras*, American Mathematical Society, Providence, R.I., 1972, Memoirs of the American Mathematical Society, No. 127.

[118] Vadim Kaimanovich, *"Münchhausen trick" and amenability of self-similar groups*, Internat. J. Algebra Comput. **15** (2005), no. 5-6, 907–937.

[119] Vadim Kaimanovich, Ilya Kapovich, and Paul Schupp, *The subadditive ergodic theorem and generic stretching factors for free group automorphisms*, Israel J. Math. **157** (2007), 1–46.

[120] Soji Kaneyuki, *On the causal structures of the Šilov boundaries of symmetric bounded domains*, Prospects in complex geometry (Katata and Kyoto, 1989), Lecture Notes in Math., vol. 1468, Springer, Berlin, 1991, pp. 127–159.

[121] Luise-Charlotte Kappe and Robert Fitzgerald Morse, *On commutators in groups*, Groups St. Andrews 2005. Vol. 2, London Math. Soc. Lecture Note Ser., vol. 340, Cambridge Univ. Press, Cambridge, 2007, pp. 531–558.

[122] Narendra Karmarkar, *A new polynomial-time algorithm for linear programming*, Combinatorica **4** (1984), no. 4, 373–395.

[123] David A. Každan and Gregory A. Margulis, *A proof of Selberg's hypothesis*, Mat. Sb. (N.S.) **75 (117)** (1968), 163–168.

[124] Ruth Kellerhals, *Collars in* $\mathrm{PSL}(2, \mathbf{H})$, Ann. Acad. Sci. Fenn. Math. **26** (2001), no. 1, 51–72.

[125] ———, *Quaternions and some global properties of hyperbolic 5-manifolds*, Canad. J. Math. **55** (2003), no. 5, 1080–1099.

[126] ———, *On the structure of hyperbolic manifolds*, Israel J. Math. **143** (2004), 361–379.

[127] John G. Kemeny and J. Laurie Snell, *Finite Markov chains*, The University Series in Undergraduate Mathematics, D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London-New York, 1960.

[128] Masashi Kiyomi, `exlp`, computer program; available from `http://members.jcom.home.ne.jp/masashi777/exlp.html`.

[129] Mustafa Korkmaz, *Stable commutator length of a Dehn twist*, Michigan Math. J. **52** (2004), no. 1, 23–31.

[130] Dieter Kotschick, *Signatures, monopoles and mapping class groups*, Math. Res. Lett. **5** (1998), no. 1-2, 227–234.

[131] ———, *Quasi-homomorphisms and stable lengths in mapping class groups*, Proc. Amer. Math. Soc. **132** (2004), no. 11, 3167–3175 (electronic).

[132] ———, *Stable length in stable groups*, Groups of Diffeomorphisms: In Honor of Shigeyuki Morita on the Occasion of His 60th Birthday, Adv. Stud. Pure Math., vol. 52, Kinokuniya, Tokyo, 2008, pp. 401–414.

[133] Frédéric Le Roux, *Simplicity of the group of compactly supported area preserving homeomorphisms of the open disc and fragmentation of symplectic diffeomorphisms*, 2009, arXiv: 0901.2428.
[134] William Bernard Raymond Lickorish, *A finite set of generators for the homeotopy group of a 2-manifold*, Proc. Cambridge Philos. Soc. **60** (1964), 769–778.
[135] Martin Liebeck, Eamonn O'Brian, Aner Shalev, and Pham Tiep, *The Ore conjecture*, submitted preprint, 2008.
[136] Bernhard Liehl, *Beschränkte Wortlänge in* SL_2, Math. Z. **186** (1984), no. 4, 509–524.
[137] Isabelle Liousse, *Rotation numbers in Thompson-Stein groups and applications*, Geom. Dedicata **131** (2008), 49–71.
[138] Roger C. Lyndon, *On Dehn's algorithm*, Math. Ann. **166** (1966), 208–228.
[139] Wilhelm Magnus, Abraham Karrass, and Donald Solitar, *Combinatorial group theory: Presentations of groups in terms of generators and relations*, Interscience Publishers [John Wiley & Sons, Inc.], New York-London-Sydney, 1966.
[140] Andrew Makhorin, `glpsol`, available from `http://www.gnu.org`.
[141] Anatoliĭ Malcev, *On isomorphic matrix representations of infinite groups*, Rec. Math. [Mat. Sbornik] N.S. **8 (50)** (1940), 405–422.
[142] ———, *On homomorphisms onto finite groups*, Ivanov. Gos. Ped. Inst. Uč. Zap. **18** (1958), 49–60.
[143] Jason F. Manning, *Quasi-actions on trees and property (QFA)*, J. London Math. Soc. (2) **73** (2006), no. 1, 84–108, With an appendix by N. Monod and B. Rémy.
[144] ———, *Virtually geometric words and Whitehead's algorithm*, arXiv: 0904.0724, 2009.
[145] Gregory Margulis, *Free subgroups of the homeomorphism group of the circle*, C. R. Acad. Sci. Paris Sér. I Math. **331** (2000), no. 9, 669–674.
[146] Andrei Markov, *Extension of limit theorems of probability theory to a sum of variables connected in a chain*, Appendix B in *Dynamic Probabilistic Systems, volume 1: Markov Chains* by Richard A. Howard, John Wiley and Sons, 1971.
[147] Bernard Maskit, *Kleinian groups*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 287, Springer-Verlag, Berlin, 1988.
[148] Howard A. Masur and Yair N. Minsky, *Geometry of the complex of curves. I. Hyperbolicity*, Invent. Math. **138** (1999), no. 1, 103–149.
[149] Shigenori Matsumoto, *Some remarks on foliated* S^1 *bundles*, Invent. Math. **90** (1987), no. 2, 343–358.
[150] Shigenori Matsumoto and Shigeyuki Morita, *Bounded cohomology of certain groups of homeomorphisms*, Proc. Amer. Math. Soc. **94** (1985), no. 3, 539–544.
[151] Dusa McDuff and Dietmar Salamon, *Introduction to symplectic topology*, second ed., Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1998.
[152] Robert Meyerhoff, *A lower bound for the volume of hyperbolic* 3*-manifolds*, Canad. J. Math. **39** (1987), no. 5, 1038–1056.
[153] Peter Milley, *Minimum-volume hyperbolic* 3*-manifolds*, 2008, arXiv: 0809.0346.
[154] John Milnor, *On the existence of a connection with curvature zero*, Comment. Math. Helv. **32** (1958), 215–223.
[155] ———, *Introduction to algebraic* K*-theory*, Princeton University Press, Princeton, N.J., 1971, Annals of Mathematics Studies, No. 72.
[156] Igor Mineyev, *Flows and joins of metric spaces*, Geom. Topol. **9** (2005), 403–482 (electronic).
[157] Nicolas Monod, *Continuous bounded cohomology of locally compact groups*, Lecture Notes in Mathematics, vol. 1758, Springer-Verlag, Berlin, 2001.
[158] Robert L. Moore, *On the foundations of plane analysis situs*, Trans. Amer. Math. Soc. **17** (1916), no. 2, 131–164.
[159] Dave Witte Morris, *Bounded generation of* $\mathrm{SL}(n,A)$ *(after D. Carter, G. Keller, and E. Paige)*, New York J. Math. **13** (2007), 383–421 (electronic).
[160] Lee Mosher and Ulrich Oertel, *Two-dimensional measured laminations of positive Euler characteristic*, Q. J. Math. **52** (2001), no. 2, 195–216.
[161] Walter D. Neumann and Don Zagier, *Volumes of hyperbolic three-manifolds*, Topology **24** (1985), no. 3, 307–332.
[162] Ulrich Oertel, *Homology branched surfaces: Thurston's norm on* $H_2(M^3)$, Low-dimensional topology and Kleinian groups (Coventry/Durham, 1984), London Math. Soc. Lecture Note Ser., vol. 112, Cambridge Univ. Press, Cambridge, 1986, pp. 253–272.

[163] Yann Ollivier, *A January 2005 invitation to random groups*, Ensaios Matemáticos [Mathematical Surveys], vol. 10, Sociedade Brasileira de Matemática, Rio de Janeiro, 2005.
[164] Oystein Ore, *Some remarks on commutators*, Proc. Amer. Math. Soc. **2** (1951), 307–314.
[165] Stephen M. Paneitz, *Invariant convex cones and causality in semisimple Lie algebras and groups*, J. Funct. Anal. **43** (1981), no. 3, 313–359.
[166] Jean-Claude Picaud, *Cohomologie bornée des surfaces et courants géodésiques*, Bull. Soc. Math. France **125** (1997), no. 1, 115–142.
[167] Henri Poincaré, *Mémoire sur les courbes définies par les équations différentialle*, J. de Mathématiques **7-8** (1881-2), 375–422, 251–296.
[168] Mark Pollicott and Michiko Yuri, *Dynamical systems and ergodic theory*, London Mathematical Society Student Texts, vol. 40, Cambridge University Press, Cambridge, 1998.
[169] Leonid Polterovich, *Floer homology, dynamics and groups*, Morse theoretic methods in nonlinear analysis and in symplectic topology, NATO Sci. Ser. II Math. Phys. Chem., vol. 217, Springer, Dordrecht, 2006, pp. 417–438.
[170] Leonid Polterovich and Zeev Rudnick, *Stable mixing for cat maps and quasi-morphisms of the modular group*, Ergodic Theory Dynam. Systems **24** (2004), no. 2, 609–619.
[171] Jerome Powell, *Two theorems on the mapping class group of a surface*, Proc. Amer. Math. Soc. **68** (1978), no. 3, 347–350.
[172] Pierre Py, *Quasi-morphisme de Calabi sur les surfaces de genre supérieur*, C. R. Math. Acad. Sci. Paris **341** (2005), no. 1, 29–34.
[173] ———, *Quasi-morphismes de Calabi et graphe de Reeb sur le tore*, C. R. Math. Acad. Sci. Paris **343** (2006), no. 5, 323–328.
[174] ———, *Quasi-morphismes et invariant de Calabi*, Ann. Sci. École Norm. Sup. (4) **39** (2006), no. 1, 177–195.
[175] ———, *Some remarks on area-preserving actions of lattices*, preprint; to appear in the Zimmer festschrift, 2008.
[176] Tibor Radó, *Über den Begriff der Riemannsche Fläche*, Acta Univ. Szeged **2** (1924–26), 101–121.
[177] Alexander Reznikov, *The volume and the injectivity radius of a hyperbolic manifold*, Topology **34** (1995), no. 2, 477–479.
[178] ———, *Quadratic equations in groups from the global geometry viewpoint*, Topology **36** (1997), no. 4, 849–865.
[179] Vsevolod I. Romanovskiĭ, *Discrete Markov chains*, Wolters-Noordhoff Publishing, Groningen, 1970, Translated from the Russian by E. Seneta.
[180] Walter Rudin, *Functional analysis*, second ed., International Series in Pure and Applied Mathematics, McGraw-Hill Inc., New York, 1991.
[181] David Ruelle, *Rotation numbers for diffeomorphisms and flows*, Ann. Inst. H. Poincaré Phys. Théor. **42** (1985), no. 1, 109–115.
[182] Saul Schleimer, *Notes on the complex of curves*, Notes from minicourse at Caltech; available from `http://www.math.rutgers.edu/~saulsch/Maths/notes.pdf`, 2005.
[183] Theodor Schneider, *Transzendenzuntersuchungen periodischer Funktionen I: Transzendenz von Potenzen*, J. Reine Angew. Math. **172** (1934), 65–69.
[184] Paul E. Schupp, *Quadratic equations in groups, cancellation diagrams on compact surfaces, and automorphisms of surface groups*, Word problems, II (Conf. on Decision Problems in Algebra, Oxford, 1976), Stud. Logic Foundations Math., vol. 95, North-Holland, Amsterdam, 1980, pp. 347–371.
[185] Peter Scott, *Subgroups of surface groups are almost geometric*, J. London Math. Soc. (2) **17** (1978), no. 3, 555–565.
[186] Nadrian Seeman and Neville Kallenbach, *DNA branched junctions*, Ann. rev. Biomol. Struct. (1994), no. 23, 53–86.
[187] Jean-Pierre Serre, *Trees*, Springer-Verlag, Berlin, 1980, Translated from the French by John Stillwell.
[188] Alexander I. Shtern, *Remarks on pseudocharacters and the real continuous bounded cohomology of connected locally compact groups*, Ann. Global Anal. Geom. **20** (2001), no. 3, 199–221.
[189] Daniel D. Sleator, Robert E. Tarjan, and William P. Thurston, *Rotation distance, triangulations, and hyperbolic geometry*, J. Amer. Math. Soc. **1** (1988), no. 3, 647–681.

[190] Craig C. Squier, *The Burau representation is unitary*, Proc. Amer. Math. Soc. **90** (1984), no. 2, 199–202.
[191] John R. Stallings, *Topology of finite graphs*, Invent. Math. **71** (1983), no. 3, 551–565.
[192] Melanie Stein, *Groups of piecewise linear homeomorphisms*, Trans. Amer. Math. Soc. **332** (1992), no. 2, 477–514.
[193] Dennis Sullivan, *A generalization of Milnor's inequality concerning affine foliations and affine manifolds*, Comment. Math. Helv. **51** (1976), no. 2, 183–189.
[194] Clifford H. Taubes, SW $\Rightarrow$ Gr*: from the Seiberg-Witten equations to pseudo-holomorphic curves*, J. Amer. Math. Soc. **9** (1996), no. 3, 845–918.
[195] Richard Thompson, handwritten notes, 1965.
[196] William P. Thurston, *A norm for the homology of* 3*-manifolds*, Mem. Amer. Math. Soc. **59** (1986), no. 339, i–vi and 99–130.
[197] ———, *Three-manifolds, foliations and circles I*, 1997, arXiv: math.GT/9712268.
[198] ———, *Geometry and Topology of* 3*-manifolds*, Notes from lectures at Princeton University, available from the MSRI, circa 1979, a.k.a. "Thurston's Notes".
[199] Takashi Tsuboi, *Homology of diffeomorphism groups, and foliated structures*, Sūgaku **36** (1984), no. 4, 320–343, Translated in Sugaku Expositions **3** (1990), no. 2, 145–181.
[200] Egbertus van Kampen, *On some lemmas in the theory of groups*, Amer. J. Math. **55** (1933), 261–267.
[201] Vladimir V. Vershinin, *Homology of braid groups and their generalizations*, Knot theory (Warsaw, 1995), Banach Center Publ., vol. 42, Polish Acad. Sci., Warsaw, 1998, pp. 421–446.
[202] Marie-France Vignéras, *Arithmétique des algèbres de quaternions*, Lecture Notes in Mathematics, vol. 800, Springer, Berlin, 1980.
[203] Carl M. Weinbaum, *Visualizing the word problem, with an application to sixth groups*, Pacific J. Math. **16** (1966), 557–578.
[204] John W. Wood, *Bundles with totally disconnected structure group*, Comment. Math. Helv. **46** (1971), 257–273.
[205] Dongping Zhuang, *Irrational stable commutator length in finitely presented groups*, J. Mod. Dyn. **2** (2008), no. 3, 499–507.

Index