Random rigidity in the free group

Danny Calegari

Alden Walker

We prove a rigidity theorem for the geometry of the unit ball in random subspaces of the scl norm in B_1^H of a free group. In a free group F of rank k, a random word w of length n (conditioned to lie in $[F, F]$) has $\mathrm{scl}(w) = \log(2k - 1)n/6\log(n) + o(n/\log(n))$ with high probability, and the unit ball in a subspace spanned by d random words of length $O(n)$ is C^0 close to a (suitably affinely scaled) octahedron.

A conjectural generalization to hyperbolic groups and manifolds (discussed in the appendix) would show that the length of a random geodesic in a hyperbolic manifold can be recovered from the bounded cohomology of the fundamental group.

Dedicated to the memory of Andrew Lange

1 Introduction

Mostow's Rigidity Theorem says that a homotopy equivalence between closed hyperbolic manifolds of dimension at least three is homotopic to an isometry. It follows that geometric invariants of a hyperbolic manifold have (at least in principle) a purely topological definition. This is most apparent in Gromov's famous proof [17] of the Rigidity Theorem, which proceeds by showing that an obviously topological invariant—namely the Gromov (or L^1) norm of the fundamental class in homology—is proportional to the volume in any hyperbolic metric. As observed by Thurston [28], a similar argument shows that for any locally symmetric space M modeled on a symmetric space X there is a constant $C(X)$ so that the norm of the fundamental class $\|[M]\|_1$ satisfies

$$\|[M]\|_1 = C(X) \cdot \mathrm{vol}(M)$$

However, the determination of the constant $C(X)$ in any given case is extremely difficult. Haagerup and Munkholm [19] showed for X equal to hyperbolic n–space $\mathbb{H}^n$ that $C(\mathbb{H}^n) = 1/v_n$, where v_n is the volume of the regular ideal hyperbolic n–simplex, and Bucher-Karlsson [4] showed that $C(\mathbb{H}^2 \times \mathbb{H}^2) = 3/2\pi^2$. The proofs are

very hard, and underscore the difficulty of computing the exact values of (nonzero) Gromov norms.

In this paper we prove a new kind of rigidity theorem for the 2–dimensional relative Gromov norm (or what is the same thing, the *stable commutator length* norm) in a free group F. This is a norm on a vector space $B_1^H(F)$, the homogenization of the space B_1 of real group 1–boundaries (in the bar complex). The space B_1^H is infinite dimensional, but its geometry can be probed by restricting attention to finite dimensional subspaces. Our main theorem is a rigidity result for the geometry of the unit ball in *random* finite dimensional subspaces of B_1^H (technically: in subspaces spanned by random elements of fixed length). We show that these unit balls are (suitably scaled) C^0 close to *octahedra* (ie, the unit ball in $\mathbb{R}^k$ with its usual L^1 norm). We also determine the *exact* scaling constant, and show that it has a simple expression in terms of the *growth exponent* of the free group (ie, the entropy of the Markov process that generates random reduced words). We concentrate in this paper on the case of free groups for clarity of exposition, but similar results should hold for random words in arbitrary hyperbolic groups, or random geodesics in negatively curved manifolds, with an analogous formula for the scaling constant. We explain the idea of this generalization in an appendix, but save the details for a follow-up paper.

Recall that *stable commutator length* is an algebraic stabilization of the topological notion of *filling genus*. If X is a space, and $\Gamma: \coprod_i S^1 \to X$ is a homologically trivial 1–manifold, the *filling genus* of Γ is the least genus of a surface S mapping to X whose boundary represents the homotopy class of Γ. The stable commutator length $\mathrm{scl}(\Gamma)$ is the infimum of $-\chi(S)/2n$ over all n and all surfaces S mapping to X whose boundary represents a cover $\widehat{\Gamma}$ of Γ of degree n. If G is a group and X is a space with $\pi_1(X) = G$, loops in X correspond to conjugacy classes in G, and the geometric definition given above defines in a natural way a *pseudo-norm* on $B_1(G)$, the space of (real) 1–boundaries; ie, finite formal real linear combinations of elements in G representing 0 in (real) homology. For G a hyperbolic group, scl descends to a norm on a suitable homogenized quotient $B_1^H(G) := B_1 / \langle g - hgh^{-1}, g^n - ng \rangle$. Precise definitions are given in Section 3.

Our first main theorem concerns the stable commutator length of a random element of $[F, F]$ of prescribed length n (we assume without comment that n is even, since a reduced element of odd length is never in $[F, F]$). Here "random" means with respect to the uniform probability on the finite set of reduced words of length n in $[F, F]$ (when n is even). For clarity, we frequently use the standard Landau "big O/little o" notation, so the expression $O(g(x))$ denotes some function $f(x)$ satisfying $f(x) \le C|g(x)|$ for some positive constant C and for all $x \gg 0$, the expression $\Theta(g(x))$ denotes some function $f(x)$ satisfying $C_1 g(x) \le f(x) \le C_2 g(x)$ for some positive constants C_1, C_2

and for all $x \gg 0$, the expression $o(g(x))$ denotes some function $f(x)$ satisfying $\lim_{x \to \infty} f(x)/g(x) = 0$, and so on. See eg [21] for a reference.

Random Rigidity Theorem 4.1 *Let F be a free group of rank k, and let v be a random reduced element of length n, conditioned to lie in the commutator subgroup $[F, F]$. Then for any $\epsilon > 0$ and $C > 1$,*

$$|\operatorname{scl}(v) \log(n)/n - \log(2k-1)/6| \le \epsilon$$

with probability $1 - O(n^{-C})$.

In particular, this implies that $\operatorname{scl}(v) \log(n)/n$ converges in probability to $\log(2k-1)/6$ as $n \to \infty$.

In more geometric language, we derive strong control on the geometry of the unit ball in the scl norm in a random subspace.

Random Norm Theorem 4.16 *Let F be a free group of rank k, and for fixed d, let $v_1, v_2, \ldots, v_d$ be independent random reduced elements of length $n_1, n_2, \ldots, n_d$ conditioned to lie in $[F, F]$, where without loss of generality we assume $n_1 \ge n_i$ for all i. Let V be the subspace of $B_1^H(F)$ spanned by the v_i. Then for any $\epsilon > 0, C > 1$ and real numbers t_i,*

$$\left|\operatorname{scl}\left(\sum t_i v_i\right) \log(n_1)/n_1 - \log(2k-1)\left(\sum |t_i| n_i\right)/6n_1\right| \le \epsilon$$

with probability $1 - O(n_1^{-C})$.

In words: the unit ball in the scl norm scaled by $n_1/\log(n_1)$ converges to the unit ball in the norm $\|\sum t_i v_i\| = \sum |t_i| n_i / n_1$ in the C^0 topology and in probability, as $n_1 \to \infty$. If $n_i = n_1 + o(n_1)$ for all i, the unit ball is C^0 close to a (scaled) octahedron.

It is worth remarking that the speed of convergence is very slow. Our asymptotic theorems depend on the distribution of the subwords of a random word at a particular characteristic scale: for a word of length n, we focus on the subwords of length $O(\log(n))$. There are some "boundary effects" that suggest a heuristic correction to our asymptotic formula, which becomes insignificant only when $\log(n)$ is sufficiently large. Computer experiments (described in Section 6) show this heuristic correction to be in very good agreement with reality. However we are not able to rigorously justify this observation nor obtain a precise asymptotic estimate of the error.

Acknowledgments

2 The random reduced word

2.1 Reduced words

Fix a free group F of rank k and a free generating set. The generators will be denoted a, b, c and so on, and their inverses by A, B, C.

We are interested in random reduced words conditioned to lie in the commutator subgroup. This is a complicated (non-local) condition to impose on a word. Fortunately, there is a nice estimate, due to Sharp, of the relative proportion of words of length n in $[F, F]$.

Theorem 2.1 (Sharp [26, Theorem 1]) *Let F be a free group of rank $k \geq 2$. Let F_n denote the set of elements of F of length n, and let $F'_n = F_n \cap [F, F]$. If n is odd, F'_n is empty, whereas there is an explicit constant σ depending on k so that*

$$\lim_{\substack{n \to \infty, \\ n \text{ even}}} \left| \sigma^k n^{k/2} \frac{|F'_n|}{|F_n|} - \frac{2}{(2\pi)^{k/2}} \right| = 0,$$

where the limit is taken over even positive integers n.

This theorem has the following consequence. Suppose that a random element of F_n has some property P with probability $1 - o(n^{-k/2})$. Then a random element of F'_n has property P with probability $1 - o(1)$. In practice, we are interested in properties of random elements in F_n that hold with probability $1 - O(C^{-n^c})$ for some constants $C > 1, c > 0$, or with probability $1 - O(n^{-C})$ for all $C > 0$, and Sharp's theorem is the fundamental tool that lets us draw conclusions about random elements of F'_n.

In the sequel we use the following notation consistently, where possible. We let v denote a random reduced word of length n, and let $m(n, k)$ (or just m for brevity) be defined by $m(n, k) := \log(n) / \log(2k - 1)$. There is a stationary Markov process that produces random reduced words in F with the uniform probability, and $\log(2k - 1)$ is the entropy of this process.

2.2 Phase transition

The constant $m = \log(n)/\log(2k-1)$ is a *natural length scale* on which to view subwords of a random word of length n. A random word of length 100 like

bbbbaBAbAABaBaabbabbaBAABBAABBAbabAAbbABBBAbaaaaBAAbbABaBabaBaBAbAABBBBaBabbaaBAAABaBabAbABaaaabbbAA

does not look homogeneous to the naked eye; the long strings of capital letters leap out and draw the reader's attention to specific locations in the word. The meaning of the scale m is that a random word of length n (for sufficiently large n) looks *homogeneous* on scales smaller than m, and *heterogeneous* on scales larger than m. However for this phase transition to become truly apparent, one must take n very large, so that $m \sim \log(n) \gg 1$.

One way to quantify this distinction is to fix a length ℓ and compute some statistic associated to the set of subwords of v of length ℓ. Each subword is an element of F_ℓ (the set of elements of F of length ℓ), and a natural number to count is

$$A_\ell(v) := \frac{1}{2n} \sum_{w \in F_\ell} \big| \text{copies of } w \text{ in } v - \text{copies of } w^{-1} \text{ in } v \big|.$$

If v is cyclically reduced, and one counts copies in the cyclic word v, then $A_\ell(v) \in [0, 1]$ with $A_\ell(v) = 1$ if and only if no inverse pair of subwords of length ℓ appear. There is a *phase transition* in A_ℓ: for $\ell = Lm$ for some fixed $L < 1$ we have $A_\ell(v) \to 0$ in probability, whereas for $\ell = Lm$ for some fixed $L > 1$ we have $A_\ell(v) \to 1$. This is proved in Sections 2.4–2.5.

For words of length $n = 10000$ in rank $k = 2$ we have $\log(n)/\log(2k-1) \approx 8.383613$. We compute $A_\ell(v)$ for a random word v in $[F, F]$ of length 10000 for $1 \le \ell \le 11$ (there are 236196 reduced words of length 11). This data is presented in Figure 1. Note that conditioning v to lie in $[F, F]$ forces $A_1(v) = 0$. The figure hints at a phase transition at $\ell \sim m$ but for it to be really sharp, one would need to take something like $n \sim$ googol.
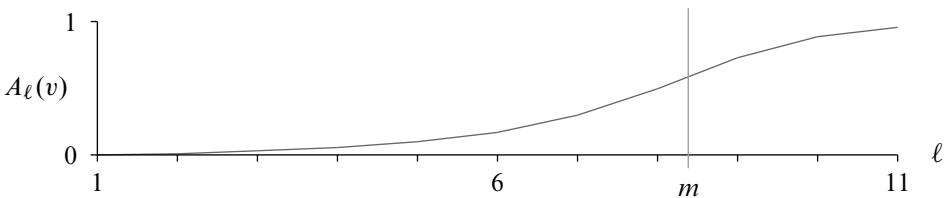

Figure 1: Values of $A_\ell(v)$ for v a random word of length 10000 and $1 \le \ell \le 12$

2.3 Counting functions and counting measures

We use the notation F_i, $F_{<i}$, $F_{\geq i}$ and so on for the set of elements in F of length i, less than i, at least i, respectively. A *random word* of length n is an element of F_n, chosen with the uniform probability measure. Note that the cardinality of F_n is $(2k)(2k-1)^{n-1}$, so $|F_{Lm}| \sim n^L$.

Although it does not add much technically, we think of F as a measure space, with the Borel algebra consisting of all subsets. Consequently any function f on F is measurable, and a function f is in $L^1(F)$ if and only if $\sum_{g \in F} |f(g)| < \infty$.

Definition 2.2 For a reduced word σ, the *counting function* C_σ is defined by

$$C_\sigma(v) = \text{number of copies of } \sigma \text{ in } v$$

and the *counting measure* $C(v)$ is the measure on F of total mass $|v|(|v|-1)/2$ for which $C(v)(\sigma) = C_\sigma(v)$.

For f a measurable function on F, define

$$C_f(v) = \int_F f \, dC(v)$$

and define $H_f(v) := C_f(v) - C_f(v^{-1})$.

2.4 Accurately estimating $C_\sigma(v)$

If v is a random word of length n, and σ is a random word of length Lm where $L < 1$, we need to estimate $C_\sigma(v)$. Since v contains $n - |\sigma| + 1$ subwords of length $|\sigma|$, the "expected" number of copies of σ in v is $(n - |\sigma| + 1)/|F_{|\sigma|}| = n^{1-L}(2k)/(2k-1) \pm O(\log(n))$. If subwords were independent, one would expect the deviation from this expected value to be typically of order $n^{(1-L)/2}$, and to be of order $n^{\epsilon+(1-L)/2}$ only with exponentially vanishing probability. This is what we prove:

Proposition 2.3 *Let $L < 1$. Then for any $\epsilon > 0$ there are constants $C > 1$ and $c > 0$ so that:*

$$\mathbf{P}\big(\big|C_\sigma(v) - n/|F_{Lm}|\big| < n^{\epsilon+(1-L)/2} \text{ for all } \sigma \in F_{Lm}\big) = 1 - O(C^{-n^c})$$

Proof The strategy is as follows. We first show that for each fixed word σ of length Lm the inequality $\mathbf{P}(|C_\sigma(v) - n/|F_{|\sigma|}|| < n^{\epsilon+(1-L)/2}) = 1 - O(C^{-n^c})$ holds. Since there are only $O(n^L) < n$ words of length Lm, it will follow that the desired estimate will hold for *every* $\sigma \in F_{Lm}$ with probability $1 - O(nC^{-n^c})$. Absorbing the n factor into the constants C and c, we will be done.

Choose some constant N (we will decide on the exact value of N later). For each residue class j mod Nm, let $v_{j,i}$ be the subword of v of length Lm that starts at the $j + iNm^{\text{th}}$ letter of v. The point is that for fixed j, the $v_{j,i}$ for consecutive i are "almost" independent. This is made precise in the following lemma:

Lemma 2.4 *For any two words x, y of length $|\sigma|$, there is an equality:*

$$\left| \mathbf{P}(v_{j,i} = x \mid v_{j,i-1} = y) - 1/|F_{|\sigma|}| \right| \leq (2k-2)^{-(N-1)m}.$$

Proof Let yuz be the subword of v starting at y, where u has length $(N-1)m$. The number of words u of fixed length for which yuz is reduced depends only on the length of u, the last letter of y, and the first letter of z. For any single letters a, b we let $u_m(a, b)$ denote the number of reduced words of the form aub of length $m + 2$. We show by induction on m that the following two statements are true:

(1) $u_m(a, b) = u_m(a, c)$ if neither of b, c are equal to $a^{-1^{m+1}}$.

(2) $|u_m(a, a^{-1^{m+1}})/u_m(a, b) - 1| \leq (2k-2)^{-m}$.

Since $u_1(a, b) = (2k-2)$ if $b \neq a$ and $u_1(a, a) = (2k-1)$ this is true for $m = 1$.

Assume it is true for $(m-1)$ odd (for example). Then depending on the first letter of u we have two cases (by the induction step), and we deduce

$$u_m(a, b) = u_{m-1}(b, b) + (2k-2)u_{m-1}(c, b) \quad \text{for } b \neq A, c \neq b,$$
$$u_m(a, A) = (2k-1)u_{m-1}(c, A) \quad \text{for } c \neq A,$$

and the induction step is proved. The case $(m-1)$ even is analogous. The lemma follows. $\qquad\square$

We resume the proof of [Proposition 2.3](). By [Lemma 2.4](), the probability that $v_{j,i} = \sigma$ conditioned on the value of $v_{j,i-1}$ is very nearly independent of the value of $v_{j,i-1}$, so we can compare the number of σ's among the $v_{j,i}$ (for fixed j) with a sum of independent Bernoulli variables, and estimate the deviation from the mean using the Chernoff bound. Let $C_{\sigma,j}(v)$ be the number of copies of σ among the $v_{j,i}$.

Lemma 2.5 *Suppose $N \geq 3$. For each j, and for any positive ϵ, there is an inequality:*

$$\mathbf{P}\left(\left| C_{\sigma,j}(v) - n/(Nm \cdot |F_{|\sigma|}|) \right| > n^{\epsilon + (1-L)/2} \right) = O(C^{-n^c}).$$

Proof By [Lemma 2.4](), the conditional probability that successive $v_{j,i}$ are equal to σ is never more than $1/|F_{|\sigma|}| + (2k-2)^{-(N-1)m}$, or less than $1/|F_{|\sigma|}| - (2k-2)^{-(N-1)m}$.

So we can bound the probability of a large deviation in terms of such large deviations for sums of independent Bernoulli trials.

Since $(2k-2)^{-(N-1)m} \leq n^{-0.6(N-1)}$ (using the estimate $\log(2k-2)/\log(2k-1) > 0.6$ for $k \geq 2$), when $N \geq 3$ we have $(2k-2)^{-(N-1)m} \leq n^{-1}$.

We have the Chernoff bound (eg the upper bound in [27, Theorem 1.3.13])

$$\mathbf{P}(|S_n - np| \geq \delta np) \leq e^{-\delta^2 np/3},$$

where S_n is a sum of n independent Bernoulli random variables with parameter p. Using $p_+ = n^{-L}(2k-1)/(2k) + n^{-0.6(N-1)} < n^{-L}$, we obtain:

$$\mathbf{P}\big(C_{\sigma,j}(v) - n/(Nm \cdot |F_{|\sigma|}|) - n^{1-0.6(N-1)}/Nm \geq \delta np_+/Nm\big) \leq e^{-\delta^2 np_+/3Nm}.$$

Since $Nm = O(\log(n))$ and $n^{1-0.6(N-1)}/Nm < 1$, taking $\delta = n^{\epsilon-(1-L)/2}Nm$, this implies

$$\mathbf{P}(C_{\sigma,j}(v) - n/(Nm \cdot |F_{|\sigma|}|) \geq n^{\epsilon+(1-L)/2}) \leq O(C^{-n^c}),$$

where $C > 1$, $c > 0$ depend only on ϵ.

A similar inequality holds for $n/(Nm \cdot |F_{|\sigma|}|) - C_{\sigma,j}(v)$. $\qquad\square$

We now complete the proof of Proposition 2.3. Since j was arbitrary, it follows that *every* $C_{\sigma,j}(v)$ deviates from $n/(Nm \cdot |F_{|\sigma|}|)$ by at most $n^{\epsilon+(1-L)/2}$, with probability at least $1 - Nm \cdot O(C^{-n^c})$ which is still $1 - O(C^{-n^c})$. Hence

$$C_\sigma(v) = \sum_j C_{\sigma,j}(v)$$

deviates from $n/|F_{|\sigma|}|$ by at most $Nm \cdot n^{\epsilon+(1-L)/2} < n^{\epsilon'+(1-L)/2}$ with the same probability. The proposition follows. $\qquad\square$

In Appendix A, we compare this result with Chernoff-type inequalities for nonreversible Markov chains obtained by Lezaud, Dinwoodie and others, and interpret such bounds in terms of the Cheeger constants of certain directed graphs.

2.5 Bounding $\sum_\sigma C_\sigma(v^{-1})$

We now turn our attention to words of length greater than m. Fix some $L > 1$, and let S be the set of subwords of v of length Lm.

Proposition 2.6 *For any ϵ there are constants $C > 1$ and $c > 0$ so that*

$$\mathbf{P}\left(\sum_{\sigma \in S} C_\sigma(v^{-1}) < n^{2-L+\epsilon}\right) = 1 - O(C^{-n^c}).$$

In particular, for $\epsilon < L - 1$, with probability $1 - O(C^{-n^c})$ there is a subset S' of S with

$$\text{card}(S - S') < n^{2-L+\epsilon} = o(n/\log(n))$$

so that no element $\sigma \in S'$ appears in v^{-1}.

Remark 2.7 Note that we think of S just as a set, *not* a set with multiplicity. For applications, it will be important to show that the cardinality of S is close to n with probability $1 - O(C^{-n^c})$; we show this as Proposition 2.11.

Remark 2.8 The set of words of length Lm has cardinality of order n^L, so the subset S has measure of order n^{1-L}. If we fix in advance any subset S of F_{Lm} of measure n^{1-L}, a robust Chernoff-type bound for Markov chains due to Lezaud (see Appendix A) gives a bound on $\sum_{\sigma \in S} C_\sigma(v^{-1})$. However this estimate cannot be applied naively to our context, since S depends (very strongly) on v.

Proof It is awkward to find a purely probabilistic proof of this estimate, because overlapping subwords of v are necessarily very highly correlated. The non-probabilistic ingredient in our proof is the following simple, but important observation:

Lemma 2.9 *Let v be a reduced word. Then for any reduced word σ, no copy of σ in v can overlap a copy of σ^{-1}.*

Proof If σ overlaps σ^{-1}, then without loss of generality we can write σ as xy where $y = y^{-1}$. But this is absurd. □

Now, for each i, let v_i be the subword of v of length Lm starting at the ith letter, and let $v_{<i}$ and $v_{>i}$ denote the part of v outside v_i, so that $v = v_{<i} v_i v_{>i}$ as a reduced word. Further, let $S_{<i}$ (respectively, $S_{>i}$) denote the subset of S consisting of subwords of length Lm in $v_{<i}$ (respectively, $v_{>i}$). By Lemma 2.9,

$$\sum_{\sigma \in S} C_\sigma(v^{-1}) = \sum_i \sum_{\sigma \in S_{<i}} C_\sigma(v_i^{-1}) + \sum_i \sum_{\sigma \in S_{>i}} C_\sigma(v_i^{-1}).$$

The point is that we can bound $\sum_{\sigma \in S_i} C_\sigma(v_i^{-1})$ in probability conditioned on $v_{<i}$, *independently* of $v_{<i}$.

Lemma 2.10 *For any* ϵ,

$$\mathbf{P}\left(\sum_{\sigma \in S_{<i}} C_\sigma(v_i^{-1}) = 1 \mid v_{<i}\right) < n^{1-L+\epsilon}.$$

Proof Note that

$$\sum_{\sigma \in S_{<i}} (v_i^{-1})$$

is 1 or 0, depending on whether v_i^{-1} is in the set $S_{<i}$ or not. No matter what $v_{<i}$ is, there are $(2k-1)^{Lm} > n^{L-\epsilon}$ choices for v_i, and each occurs with the uniform probability. The cardinality of $S_{<i}$ is at most i which is less than n, so the chance that v_i^{-1} is in $S_{<i}$ is at most $n^{1-L+\epsilon}$, as claimed. $\square$

It follows that if we fix a residue $j \mod Lm$, for any ϵ there are $C > 1$, $c > 0$ such that we can estimate

$$\mathbf{P}\left(\sum_{i=j \bmod Lm} \sum_{\sigma \in S_{<i}} C_\sigma(v_i^{-1}) \geq n^{2-L+\epsilon}/Lm\right) < O(C^{-n^c}).$$

Summing over all residue classes j, and then replacing $S_{<i}$ by $S_{>i}$ by symmetry proves the proposition. $\square$

As remarked above, it is important for applications to show that the cardinality of S is very close to n, with high probability.

Proposition 2.11 *Fix $L > 1$ and let S denote the set of subwords of v of length Lm. There is an ϵ and $C > 1$, $c > 0$ so that*

$$\mathbf{P}(n - \mathrm{card}(S) > n^{1-\epsilon}) = O(C^{-n^c}).$$

Proof The proof is almost the same as that of [Proposition 2.6](), except that we need to estimate the number of σ for which some copy of σ overlaps itself, and show this is less than $n^{1-\epsilon}$ for some ϵ with the desired probability.

There are two kinds of overlaps to consider: those for which the nonoverlapping initial segment of the first word has length less than $2m/3$ ("big overlaps") and those for which it has length at least $2m/3$ ("little overlaps"). We count the number of each independently.

A big overlap results in a subword of the form wuw where the length of w is at least $m/6$ and the length of u is at least $m/6$. Conditioned on w and u, the probability that the next word will be a copy of w is at most $n^{-1/6}$, so there are at most $n^{5/6}$

subwords of v that are contained in a big overlap. A little overlap results in a subword of the form ww where the length of w is at least $2m/3$. Again, conditioned on w, the probability that the next word will be a copy of w is at most $n^{-2/3}$ so there are at most $n^{1/3}$ subwords of v that are contained in little overlaps. Each subword is contained in at most $Lm = O(\log(n))$ overlaps of either kind. The result follows. □

3 Stable commutator length

The material in this section is standard. A basic reference is Calegari [5].

3.1 Definitions

Definition 3.1 Let G be a group, and $[G, G]$ the commutator subgroup. The *commutator length* of an element $g \in [G, G]$, denoted $\mathrm{cl}(g)$, is the least number of commutators whose product is g, and the *stable commutator length*, denoted $\mathrm{scl}(g)$, is the limit $\mathrm{scl}(g) := \lim_{n \to \infty} \mathrm{cl}(g^n)/n$.

The definition of (stable) commutator length can be extended to finite formal sums as follows:

Definition 3.2 Let G be a group, and let $\{g_i\}$ be a finite collection of elements with $\prod_i g_i \in [G, G]$. Define $\mathrm{cl}(\sum g_i)$ to be the minimum of $\mathrm{cl}(\prod g_i^{h_i})$ over all products of conjugates $g_i^{h_i}$ of the g_i. This is symmetric, and a class function in each g_i separately. Define $\mathrm{scl}(\sum g_i) = \lim_{n \to \infty} \mathrm{cl}(\sum g_i^n)/n$.

Let $C_1(G)$ be the real vector space with basis the elements of G, and let $B_1(G)$ be the kernel of $C_1(G) \to H_1(G; \mathbb{R})$. So $B_1(G)$ is the space of formal finite real linear combinations of elements in G that represent 0 in (real) homology. Equivalently, $B_1(G)$ is the image of the vector space of real 2–chains (in the bar complex) under ∂. It is a fact that scl extends by linearity and continuity to a pseudo-norm on $B_1(G)$, and vanishes on the subspace $\langle g - hgh^{-1}, g^n - ng \rangle$. This vanishing reflects the homogeneity of scl and the fact that it is a class function in each variable separately. So scl descends to a pseudo-norm on the quotient $B_1^H(G) := B_1(G)/\langle g - hgh^{-1}, g^n - ng \rangle$.

The following theorem is nice to know, but is not used in an essential way in this paper:

Theorem 3.3 (Calegari and Fujiwara [7]) *Let G be (word) hyperbolic. Then* scl *is a norm on* $B_1^H(G)$.

3.2 Surfaces

Let X be a space with $\pi_1(X) = G$, and for any finite collection of conjugacy classes g_i let $\Gamma\colon \coprod_i S_i^1 \to X$ be a 1–manifold in the associated free homotopy class. A map of a (compact, oriented) surface $f\colon S \to X$ is *admissible* if there is a commutative diagram

$$
\begin{array}{ccc}
\partial S & \longrightarrow & S \\
\partial f \downarrow & & \downarrow f \\
\coprod_i S_i^1 & \xrightarrow{\ \Gamma\ } & X
\end{array}
$$

and an integer $n(S)$ for which $\partial f_*[\partial S] = n(S)[\coprod_i S_i^1]$ in H_1. The map is *monotone* if $\partial S \to \coprod_i S_i^1$ is homotopic to an orientation-preserving cover (equivalently, if every component of ∂S wraps with positive degree around its image).

Lemma 3.4 [5, Proposition 2.74] *Let $g_1, \ldots, g_m$ be conjugacy classes in G, represented by $\Gamma\colon \coprod_i S_i^1 \to X$. Then*

$$
\mathrm{scl}\left(\sum_i g_i\right) = \inf_S \frac{-\chi^-(S)}{2n(S)},
$$

where the infimum is taken over all surfaces S and all maps $f\colon S \to X$ admissible for Γ.

The notation $\chi^-(S)$ means the sum of Euler characteristics $\sum_i \chi(S_i)$ taken over those components S_i of S with $\chi(S_i) \le 0$. By [5, Proposition 2.13] it suffices to restrict to *monotone* admissible surfaces. An admissible surface S is *extremal* if equality is achieved.

3.3 Fatgraphs

If F is free, X can be taken to be a graph, and any admissible surface can be represented combinatorially (possibly after performing some compressions) by a *fatgraph*. Fatgraphs are combinatorial objects which allow one to move back and forth between group theory/combinatorics and 2–dimensional topology; a standard reference is Penner [23], especially Section 1.

A *fatgraph* Y is a graph together with a cyclic ordering of the edges incident at each vertex. Such a graph can be thickened to a compact surface $S(Y)$ (or just S if Y is understood) in such a way that Y embeds in $S(Y)$ as a deformation retract. A

fatgraph Y is *oriented* if $S(Y)$ is oriented. In the sequel we assume all our fatgraphs are oriented, and have no 1–valent vertices. Note that $\chi(Y) = \chi(S(Y))$.

A *fatgraph over F* is a fatgraph with oriented edges labeled by words in F so that opposite sides get inverse labels, and the cyclic words obtained by reading around $\partial S(Y)$ are reduced. By abuse of notation we write ∂Y in place of $\partial S(Y)$ and think of it as an element of $B_1^H(F)$. Figure 2 gives an example of an extremal fatgraph for the chain $a + b + AB + [a, b]$ in F_2. Note that extremal surfaces do not need to be connected.
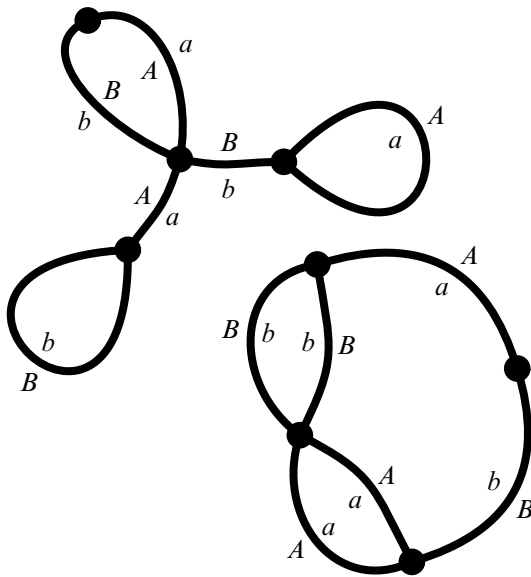

Figure 2: An extremal surface for $a + b + AB + [a, b]$ represented as a fatgraph

The basic fact we use is the following lemma, which is a restatement of [14, Theorem 1.4] in the language of fatgraphs.

Lemma 3.5 (Culler [14, Theorem 1.4] (fatgraph lemma)) *Let S be an admissible surface bounding a chain Γ. Then after possibly compressing S a finite number of times (thereby reducing $-\chi^-(S)$ without changing ∂S) there is a fatgraph Y over F with $S(Y) = S$ and $\partial Y = \Gamma$.*

Remark 3.6 Culler proves his theorem only for surfaces with connected boundary, but his argument generalizes with no extra work. An equivalent statement, valid for surfaces with disconnected boundary, is also proved in Calegari [6, Lemma 3.4]; also see [5, Section 4.3] for a discussion and references.

Let Y be an extremal fatgraph for v. The underlying fatgraph might not be trivalent, but by splitting higher valence vertices, and inserting (unlabeled) "dummy edges", we can think of Y as a trivalent fatgraph in a degenerate way, where some degenerate "edges" have length 0. We call this the operation of *resolving vertices* (such a resolution need not be unique).

Lemma 3.7 *Let Y be an extremal fatgraph for v, so that ∂Y represents Nv for some N, and $-\chi(Y)/2N = \mathrm{scl}(v)$. Resolve vertices of Y so that Y is trivalent, possibly with some edges of length 0. Let the average length of the edges of Y be ℓm. Then*

$$\mathrm{scl}(v) = n \log(2k - 1)/12\ell \log(n).$$

Proof Suppose Y has V vertices and E edges. Since Y is trivalent, $2E/3 = V$ and $-\chi(Y) = E - V = E/3$. On the other hand, the total length of ∂Y is $Nn = 2E\ell m$. Hence

$$\mathrm{scl}(v) = -\chi(Y)/2N = E/6N = n/12\ell m = n \log(2k - 1)/12\ell \log(n). \qquad \square$$

It will be our goal to show that for *random* v of length $n \gg 1$, the extremal fatgraph Y has $\ell = 1/2 + o(1)$ with probability $1 - o(1)$.

Remark 3.8 The reader who is unhappy with edges of length 0 can just take ℓm to be equal to the total length of Y divided by $E + \sum_v \mathrm{valence}(v) - 3$.

4 Random values of scl

The goal of this section is to prove the Random Rigidity Theorem:

Theorem 4.1 (Random Rigidity Theorem) *Let F be a free group of rank k, and let v be a random reduced element of length n, conditioned to lie in the commutator subgroup $[F, F]$. Then for any $\epsilon > 0$ and $C > 1$,*

$$|\mathrm{scl}(v) \log(n)/n - \log(2k - 1)/6| \le \epsilon$$

with probability $1 - O(n^{-C})$.

The proof will occupy most of the remainder of the section.

4.1 Upper bounds

The upper bound in the Random Rigidity Theorem is sharpened by the following proposition:

Proposition 4.2 *Let v be a random reduced word in the commutator subgroup of length n. Then for any $\epsilon > 0$ there are constants $C > 1$ and $c > 0$ so that*

$$\mathrm{scl}(v)\log(n)/n - \log(2k-1)/6 \leq \epsilon$$

with probability $1 - O(C^{-n^c})$.

Given random v, we explicitly build an extremal surface (actually an extremal fatgraph) by gluing together a very large number of *tripods* with edges of length slightly less than $(1-\epsilon)m/2$. The fact that such tripods can be glued up to produce a fatgraph with boundary very close to a multiple of v follows from an equidistribution lemma, derived from the estimates in Section 2, which holds with very high probability for most random v. The tripods do not glue up completely, but the mass of the unglued part has size $O(n^{-\epsilon/2})$ compared to the glued part, and the remainder can be glued up (under the hypothesis that v is homologically trivial) with a contribution to χ proportional to the mass.

4.2 Tripods and joints

In what follows we generally adhere to the notational convention that group inverses are denoted by small and capital letters; hence X means x^{-1} and so on.

Definition 4.3 A *tripod of edge length L* is a fatgraph with underlying graph a tripod, and with edges labeled by reduced words xY, yZ, zX where each of x, y, z (the *incoming edge labels*) has length L. We denote such a tripod $T(x, y, z)$.

A *copy* of $T(x, y, z)$ is a triple of segments of the form xY, yZ, zX in v. These segments may appear anywhere in v; they might or might not be adjacent, and are allowed to overlap each other.

Lemma 4.4 *A triple x, y, z of reduced words of length L are the labels of a tripod if and only if their last letters are distinct. Consequently, for any reduced word xY of length $2L$, there are $(2k-2)(2k-1)^{L-1}$ choices for z.*

Proof Obvious. □

There are $(2k)(2k-1)(2k-2)(2k-1)^{3(L-1)}/3 \sim (2k-1)^{3L}/3$ tripods T of edge length L. For each tripod T, let ∂T denote the triple of words xY, yZ, zX.

Definition 4.5 A *joint of edge length* L is a fatgraph with underlying graph a segment, and with edges labeled by reduced words x, X each of length L. Denote such a joint $J(x)$.

A *copy* of $J(x)$ is an *ordered* pair of segments of the form x, X in v. Again, these segments may appear anywhere in v (note that since v is reduced, these segments cannot overlap or be adjacent in v). We distinguish between orientations, so that $J(x)$ and $J(X)$ are different.

Each joint $J(x)$ is contained in a unique maximal joint $J(x')$.

Fix L with $L/m = 1/2 - \epsilon$ for some small ϵ. For a word v, let $T_L(v)$ denote the set of copies of tripods of edge length L in v, and let $J_L(v)$ denote the set of copies of joints of edge length L in v. Note that each pair of subwords x, X of length L in v determines *two* elements of $J_L(v)$. We define an involution ι on the set $J_L(v)$ interchanging such pairs. If v is understood, we just write T_L and J_L.

Given T, a copy of $T(x, y, z)$ of length L, there are three associated joints $J(x)$, $J(y)$, $J(z)$ which can be extended uniquely to maximal joints $J(x')$, $J(y')$ and $J(z')$. Note that x is a suffix of x', and so on. Define $\partial T(x, y, z) = J(x') + J(y') + J(z')$ and extend ∂ to a linear map from the space of measures on T_L to the space of measures on J_L.

Example 4.6 Let $v = ABBAbAABAAbababaabbABBBBabbaaB$. The tripod of length 2 as indicated

$$ABB\underline{Ab}AA BAAbabab\underline{aabb}AB\underline{BBB}abbaaB$$

is associated to three joints: a pair Ab, Ba, a pair AA, aa and a pair bb, BB. The joint Ab, Ba is contained in a maximal joint of length 5

$$\underline{ABBAb}AABAAbababaabbABBB\underline{Babbaa}B$$

and the joint aa, AA is contained in a maximal joint of length 4

$$ABBAb\underline{AABA}Abab\underline{abaabb}ABBBBabbaaB$$

whereas the joint bb, BB of length 2 is already maximal:

$$ABBAbAABAAbababa\underline{abb}AB\underline{BB}BabbaaB$$

The next lemma, although a simple consequence of the estimates in Section 2.4, is key. It shows that with very high probability, the collection of *all* tripods of length $(1/2 - \epsilon)m$ can be almost exactly glued up in pairs:

Lemma 4.7 *Let $L/m = 1/2 - \epsilon$. Then with probability $1 - O(C^{-n^c})$ there is an inequality $|\partial \mu - \iota \partial \mu| = O(n^{-\epsilon/3}|\mu|)$, where μ is the uniform measure on T_L, and $|\cdot|$ denotes mass of a (possibly signed) measure.*

Proof For any given $J(x)$ in v contained in a maximal $J(x')$ we estimate the number of tripods $T(x', y, z)$ with $J(x')$ in ∂T. First of all, y is determined, since the copy of x associated to J is the initial subword of some xY. Similarly, z is determined, since the copy of X associated to J is the terminal subword of some zX. Therefore the number of tripods is simply equal to the number of subwords of the form yZ in v.

The number of copies of yZ in v is approximately $n/|F_{2L}|$, ie, about n^ϵ with an error of size $O(n^{\epsilon/2+\delta})$ for any δ, by Proposition 2.3. Taking $\delta = \epsilon/6$ for concreteness, the error is at most $O(n^{2\epsilon/3})$ which is a fraction $O(n^{-\epsilon/3})$ of the total mass.

Since this is true for every joint $J(x)$, the lemma follows. □

4.3 Proof of upper bound

The proof of Proposition 4.2 is now straightforward:

Proof Assemble the tripods and glue them in pairs along their common boundary joints. By Lemma 4.7 all but $O(n^{-\epsilon/3})$ of the measure of the set of tripods can be glued up this way, with probability $1 - O(C^{-n^c})$. This holds even conditioning on $v \in [F, F]$ with probability $1 - O(C^{-n^c})$, with slightly different constants, by Theorem 2.1.

This (partial) fatgraph Y can be extended (usually in many ways) to a complete fatgraph bounding some multiple of v in $B_1^H(F)$ so that the Euler characteristic of the added surface is proportional to the mass of the unglued part. We explain how to do this.

Let N be the function on the letters of v whose value at a given letter is the number of edges of tripods that contain it, and let N' be the maximum of N. The function $N' - N$ is therefore non-negative, and on the other hand max $N' - N = O(n^{-\epsilon/3})N'$. We translate the problem of building a fatgraph that extends Y as a problem of suitably gluing together a collection of rectangles.

Each rectangle corresponds to some finite subword w of v which we call the *label* of the rectangle. We think of the rectangle as having height 1 and width equal to the length of w. We keep track not only of w as a word in the generators, but also of
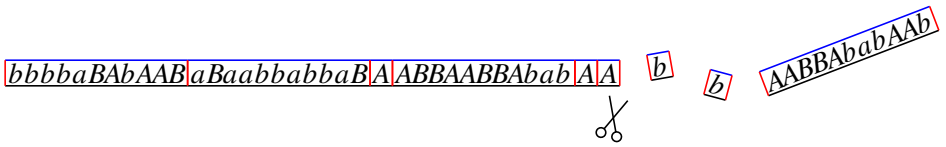
Figure 3: N' copies of v cut up into long and short rectangles

where it appears as a subword of v. Color the top horizontal edge of the rectangle blue, and the vertical sides red.

We want to glue together rectangles along segments of the boundary of integer length, blue to blue and red to red, so that two red edges may be glued only if the words associated to the rectangles are consecutive subwords of v, and two blue segments are glued only if the paired letters on either side are inverse in F.

We take three rectangles for each copy of each tripod, with labels the subwords of v corresponding to the edges of the tripod. We also take $N' - N$ rectangles for each letter of v, with label that letter. So we have lots of "long" rectangles — three for each tripod — and far fewer "short" rectangles (of length 1). By the definition of N' and N, every letter of v appears as the rightmost letter of a label exactly as many times as the following letter of v appears as the leftmost letter of a label. So we could think of taking N' strips labeled v and cutting them into long and short rectangles; see Figure 3. Naturally, it is possible to glue up the red segments in pairs compatibly. However, there are potentially many ways to do this, and it is important to glue up blue edges first, as we now explain.

The long rectangles can be glued up along blue edges in threes to build fattened tripods. Pairs of tripods can then be glued up along red edges corresponding to joints (note that pairs of tripods are glued up in this manner along red segments of length 2). The result can be thought of in an obvious way as the partial fatgraph Y, where the blue edges are the core graph. See Figure 4.

Recall that by hypothesis v is homologically trivial, and note that the rectangles corresponding to a given tripod have the same number of copies of each generator as of its inverse. Consequently for each generator of F, there are as many short rectangles labeled with this generator as are labeled with its inverse. We can therefore glue together these short rectangles in pairs, so that every blue edge can be thus glued up.

As observed above, the remaining unglued red segments can be glued up in pairs. We now perform this gluing (in an arbitrary way). See Figure 5. Note that the result might have corners at which more than two paired red edges meet.
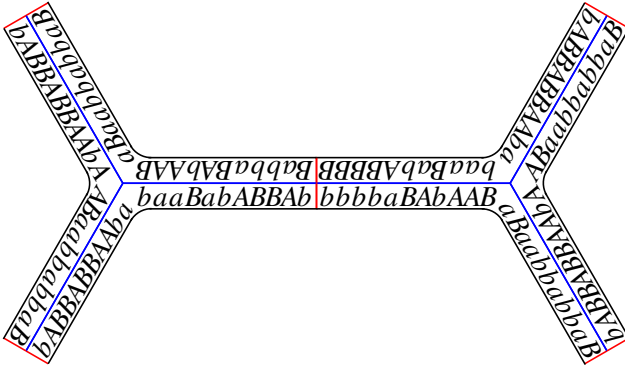
Figure 4: Long rectangles glued in threes along blue edges to make tripods, and tripods glued in pairs along red edges
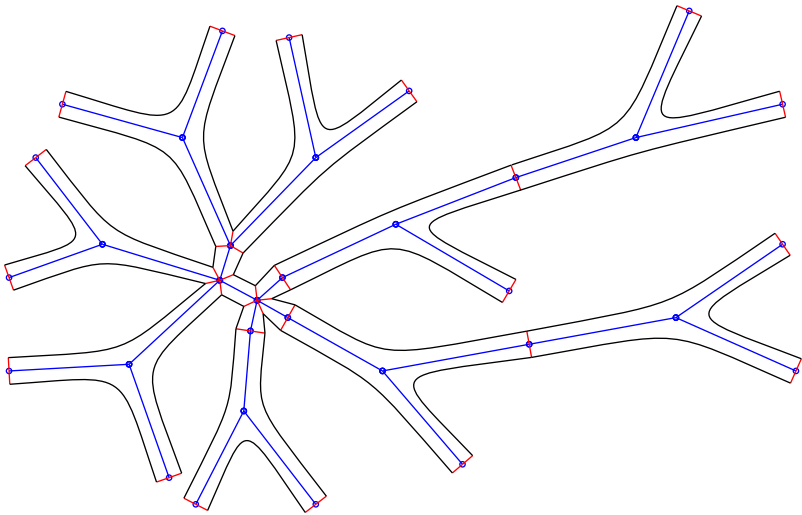

Figure 5: Remaining red segments can be glued in pairs. This might produce red corners where more than two red edges meet.

The resulting surface Y' has no unglued red edges. The blue edges form the core of the surface, and the labels and the way the blue edges sit in the surface amounts to giving it the structure of a fatgraph over F whose boundary is a multiple of v. The fatgraph Y sits in Y' in an obvious way, and the contribution of $Y' - Y$ to $-\chi$ is of order $(N' - N)|v|$, which is very small compared to the contribution from Y. In particular, the average edge length ℓm of Y' differs from the average edge length of

Y by at most $O(n^{-\epsilon/3})$, and therefore satisfies $\ell > 1/2 - \epsilon$. The proof now follows from Lemma 3.7. □

Remark 4.8 The use of ergodic theory to construct an almost equidistributed collection of pieces with prescribed geometry that can be almost glued up is inspired by the techniques in Kahn and Markovic's recent proof [20] of the surface subgroup conjecture in 3–manifold topology, and we are pleased to acknowledge our intellectual debt to this paper.

4.4 Lower bounds

The goal of the next few sections is to prove the following estimate, which precisely complements Proposition 4.2. The Random Rigidity Theorem (ie, Theorem 4.1) follows immediately from these two propositions.

Proposition 4.9 *Let v be a random reduced word in the commutator subgroup of length n. Then for any $\epsilon > 0$ and any C,*

$$\log(2k-1)/6 - \mathrm{scl}(v)\log(n)/n \le \epsilon$$

with probability $1 - O(n^{-C})$.

Note that the probability estimate associated to the upper bound is exponential, whereas the estimate associated to the lower bound is merely polynomial (of arbitrarily large degree). This disparity is an artifact of the method of proof. A worse lower bound, but with exponential bounds on the probability of deviation, is obtained in Section 5 using the method of *quasimorphisms*.

4.5 Combs

Let b be a subword of v, and consider some copy of b in the boundary of an extremal fatgraph Y for v. Recall that by our convention we artificially split open vertices of higher valence so that Y is trivalent, although it might have some edges of length 0. The subword b is contained in a segment σ of Y, which is incident to a sequence of edges $e_1, e_2, \ldots, e_d$ of Y in order. Call the subgraph of Y consisting of the support of b together with the union of the e_i a *comb*.

Let $c_1, \ldots, c_d$ be the labels on the edges e_i (oriented to point in to σ). Furthermore, the vertices of the e_i subdivide b into subwords $b_0, \ldots, b_d$, where we stress that some b_i, c_i might have length 0. Then there are boundary labels of Y of the form $B_d C_d, c_d B_{d-1} C_{d-1}, \ldots, c_2 B_1 C_1, c_1 B_0$ (see Figure 6). By the definition of an extremal fatgraph, these boundary labels are (cyclic) subwords of v.
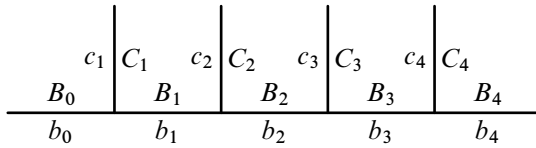
This suggests the following definition:

Figure 6: A comb with edge labels

Definition 4.10 Given a word $b \subset v$ a *comb on* b is a family of subwords of v of the form $B_d C_d, c_d B_{d-1} C_{d-1}, \ldots, c_1 B_0$. The *complexity* of the cone is d (as above) and the *length* is L, where $|b| + \sum |c_i| = Lm$.

We would like to bound (in probability) the length of a comb in terms of its complexity. Fix a big constant L', and let $b \subset v$ be a subword of length $L'm$. We would like to construct a comb on b for which L/d is as big as possible. This amounts to choosing a partition of b into d successive subwords b_i of length $L_i m$ (where $L_i = 0$ is allowed), then choosing copies of B_i in v, and defining c_i to be the maximal subword following the copy of B_i for which C_i precedes the copy of B_{i-1}. Let these maximal c_i have length $K_i m$.

Note that the comb has length $L = \sum_{i=0}^{d} L_i + \sum_{i=1}^{d} K_i$ and complexity d. We would like to bound in probability the maximum ratio $L/(2d+1)$, at least for typical b of some fixed length $L'm$ where $L' = \sum L_i$.

By Propositions 2.3 and 2.6 there are almost exactly n^{1-L_i} possible locations of each B_i in v for $L_i < 1$, and the chance that there is some B_i at all when $L_i > 1$ is at most n^{1-L_i}. If we assume that the prefixes and suffixes of the B_i of fixed length are evenly distributed, then for any fixed T, there should be an estimate

$$\mathbf{P}\left(\sum K_i > T + \sum(1 - L_i)\right) = O(n^{-T}).$$

If T is very big but fixed, and small compared to $L' = \sum L_i$, then we can estimate $L \le T + (d+1)$, and therefore $L/(2d+1) \le 1/2 + \epsilon$ for any ϵ with probability $1 - O(n^{-T})$. This is good enough to give the desired bound in Proposition 4.9, by Lemma 3.7.

Notice that this heuristic argument is almost rigorous: prefixes and suffixes of the B_i are not perfectly independent, but their correlation decays exponentially fast with the distance between B_i and B_j. Thus we need only examine the cases in which there are $C_{i+1} B_i c_i$ and $C_{j+1} B_j c_j$ that overlap. In order to obtain the desired estimate, it is necessary to make some *a priori* assumptions about a cone on b, which will turn out to be justified for *most* combs in any given extremal fatgraph Y.

Definition 4.11 A subword b of v is δ–*regular* if there is no subword b' of length $(1+\delta)m$ such that B' is in v, and if all subwords of b of length δ and their inverses are distinct.

A comb on b is δ–*regular* if b is δ–regular, and if all the c_i have length at most $(1+\delta)m$.

Let v be a random word in F' of length n, and let Y be an extremal trivalent fatgraph for v (possibly with some edges of length 0). For any d, we can consider the set of combs of Y of complexity d. The following lemma justifies the definition of δ–regular:

Lemma 4.12 *Let v be a random word in F' of length n, and let Y be an extremal fatgraph for v. Then for any d, the proportion of combs of Y of complexity d that are not δ–regular is at most $O(n^{-\delta/2})$, with probability $1 - O(C^{-n^c})$.*

Proof By Proposition 2.6, with probability $1 - O(C^{-n^c})$ there are at most $n^{1-\delta/2}$ subwords of v of length $(1+\delta)m$ whose inverse also appears in v (in fact, we could take any number less than δ in place of $\delta/2$); hence the proportion of combs of complexity d that contain an edge of length at least $(1+\delta)m$ is at most $(4d+2)n^{-\delta/2}$, since every edge of Y is contained in $4d+2$ combs of complexity d, and ∂Y represents Nv for some N.

An argument similar to Proposition 2.11 establishes that subwords of typical b of length δ are distinct, with probability $O(C^{-n^c})$. □

4.6 Overlaps

We now restrict attention to a fixed δ–regular word b, and consider a random word v conditioned to contain b as a subword. The arguments in this section depend on order-of-magnitude estimates of probability, expressed as a power of n.

Fix vectors of lengths L_i, $K_i < (1+\delta)m$, and for each choice of d locations in v, consider the probability that the subwords $e_i D_{i-1} C_i$ of length $K_i + L_{i-1} + K_{i-1}$ starting at these locations constitute a comb on b; we call such an occurrence a *matching*, and we want to estimate the probability of a matching at a given d–tuple of locations. We also refer to a vector of d locations in v as above as a *configuration*. If the subwords do not overlap, this probability is less than $n^{-(\sum L_i + \sum K_i)}$. So it suffices to estimate the probability in the case that some subwords do overlap. This is somewhat fiddly, and depends on an analysis of the combinatorial possibilities for the overlap. However, the estimates in every case are entirely elementary.

For each $j \geq 2$, let $P_j m$ be the total length where at least j words overlap. Define the *total overlap*, counted with multiplicity, to be $P := \sum_{j \geq 2} P_j$. The total contribution to P from overlaps of D_i with D_j will be $O(\delta)$, since b is δ–regular. If part of some e_i (resp. C_i) is contained in an overlap, but the corresponding part of C_i (resp. e_i) is not, this overlap does not significantly affect the probability of a matching. If corresponding parts of C_i, e_i both overlap D_j, then again necessarily this overlap will be of size $O(\delta)m$, since b is δ–regular. So to estimate the probability of a matching, it suffices to consider overlaps among the various C_i, e_j. Let $P'_j m$ be the total length where at least j such subwords overlap, and analogously define $P' := \sum_{j \geq 2} P'_j$.

Lemma 4.13 *With notation as above, the probability of a matching in a given configuration is at most* $n^{P'/2 - (\sum L_i + \sum K_i) + O(\delta)}$.

Proof An overlap in some subword of e_i of length lm must correspond to an overlap in the corresponding subword of C_i to increase the probability of a match by at most n^l; so the increase over the "naive" probability of a match is at most a factor of $n^{P'/2}$. $\square$

On the other hand, there are n^d sets of locations of the subwords, and for each given location of one subword, there are only $O(\log(n))$ locations of any other subword that overlaps it. Two subwords $e_i D_{i-1} C_{i-1}$ and $e_j D_{j-1} C_{j-1}$ can contribute at most $2(1 + \delta)$ to P', precisely if $e_i = e_j$ and $C_{i-1} = C_{j-1}$. We deduce the following lemma:

Lemma 4.14 *Let L_i, K_i be some fixed vector of lengths with $L_i, K_i < 1 + \delta$, and define $L = \sum_i L_i + \sum_i K_i$. Suppose b is an δ–regular subword of v. Then the probability that there is a comb over b with the prescribed lengths is at most $O(n^{-T + O(\delta)})$ where $T = L - d - 1$. Consequently if $L/(2d + 1) \geq 1/2 + \epsilon$ and δ is sufficiently small compared to ϵ, and d is sufficiently big compared to ϵ, we can make T as big as desired.*

Proof As above, each set of locations has probability at most $n^{P'/2 - L + O(\delta)}$ of a matching. Moreover, there are n^d sets of locations, and at most $n^{d-r+O(\delta)}$ sets of locations for which $P' \geq 2r$. The estimate follows. $\square$

4.7 Proof of lower bound

We now give the proof of .

Proof By Lemma 3.7, it suffices to show for every C and every ϵ that the average length ℓm of the edges of an extremal fatgraph Y is at most $1/2 + \epsilon$, with probability $1 - O(n^{-C})$. By Theorem 2.1, conditioning that v lies in $[F, F]$ only affects probabilities by at most a factor of $O(n^{k/2})$.

By Proposition 2.6, there are only $O(1)$ subwords of v of length at least $2m$ and $O(n^{1-\delta/2})$ of length at least $(1+\delta)m$, whose inverse also appears in v, with probability $1 - O(C^{-n^c})$. So edges of length at least $(1+\delta)m$ affect ℓ negligibly, and the fraction of combs containing such subwords are similarly negligible.

Choose some very large constant d, roughly of size $O(1/\epsilon)$, and consider the set of all combs with complexity d in Y. Because Y is (formally) trivalent, every edge occurs in exactly $(4d + 2)$ such combs; each comb has $2d + 1$ edges, and each edge has two sides. By Lemma 4.12, if $\ell \geq 1/2 + \epsilon$, a definite fraction of these combs must be δ–regular, and satisfy $L/(2d + 1) > 1/2 + \epsilon$.

On the other hand, by Lemma 4.14, for any δ–regular subword b and any given vector of lengths less than $1 + \delta$ the probability that there is a comb over b with prescribed lengths is at most $O(n^{-T+O(\delta)})$ where $T = L - d - 1$. Since there are at most n possible locations in v for such a subword b, and since there are at most $((1+\delta)m)^{2d+1} < n^{\delta}$ vectors of lengths, the probability that there is *any* δ–regular comb with complexity d and length L is at most $O(n^{-T+1+O(\delta)})$. So for any C and any ϵ, if d is sufficiently large and $L/(2d + 1) > 1/2 + \epsilon$, *no* such comb exists, with probability $1 - O(n^{-C})$. The proof follows. $\square$

Remark 4.15 A more careful analysis would almost certainly improve the estimate of the probability of a large negative deviation. The probability that a specific δ–regular subword is part of a δ–regular comb with big d and $L/(2d+1) > 1/2+\epsilon$ is polynomial in n, and to violate the desired lower bound on scl we must construct a fatgraph containing a *definite proportion* of such big δ–regular combs. However, the events that distinct subwords b, b' are parts of such δ–regular combs are not obviously independent, and even estimating their correlation appears hard. Nevertheless, heuristically one would expect the true probability of a deviation to be exponential in (some power of) n.

4.8 The Random Norm Theorem

In fact, it is not much more work to derive the following theorem, which specializes to Theorem 4.1 when $d = 1$:

Theorem 4.16 (Random Norm Theorem) *Let F be a free group of rank k, and for fixed d, let $v_1, v_2, \ldots, v_d$ be independent random reduced elements of length*

$n_1, n_2, \ldots, n_d$ conditioned to lie in $[F, F]$, where without loss of generality we assume $n_1 \geq n_i$ for all i. Let V be the subspace of $B_1^H(F)$ spanned by the v_i. Then for any $\epsilon > 0, C > 1$ and real numbers t_i,

$$\left| \mathrm{scl}\left(\sum t_i v_i \right) \log(n_1)/n_1 - \log(2k-1)\left(\sum |t_i| n_i \right)/6n_1 \right| \leq \epsilon$$

with probability $1 - O(n_1^{-C})$.

We remark before giving the proof that even though the C^0 geometry of a (random) slice of the unit ball is very simple, the finer polyhedral structure is apparently extremely complicated. Figure 7 and Figure 8 exhibit 2 and 3 dimensional slices of the scl unit ball of some relatively simple words.
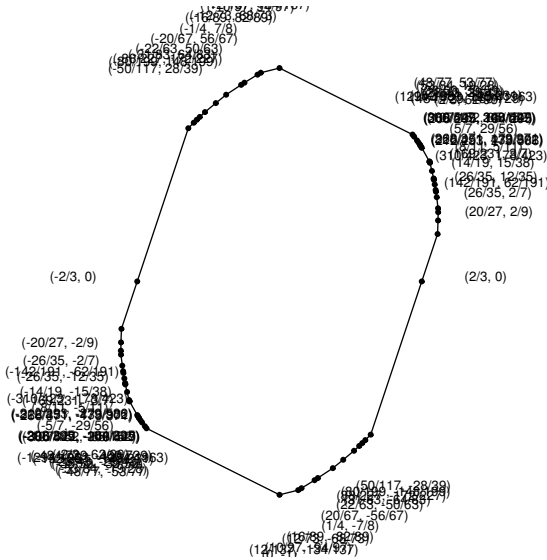

Figure 7: Unit ball in the scl norm in the subspace spanned by
$Aba BAba BBAbaabaa BAAA$ and $baaab A Abaab A BAA BBA Ba$

Proof We give the proof in the case $d = 2$; the general case follows by essentially the same argument. For any pair of reduced words v_1, v_2 (not necessarily in $[F, F]$) choose a word z of length 6 contained in $[F, F]$ so that $v_1 z v_2$ is reduced. We can always find such a word z of the form $xyXXYx$ for some generators x, y so that v_1 does not end, and v_2 does not begin, with X.

This defines a map $F_{|v_1|} \times F_{|v_2|} \to F_{|v_1|+|v_2|+6}$, and the pushforward of the product of uniform measures is proportional to the uniform measure on the image, with constant
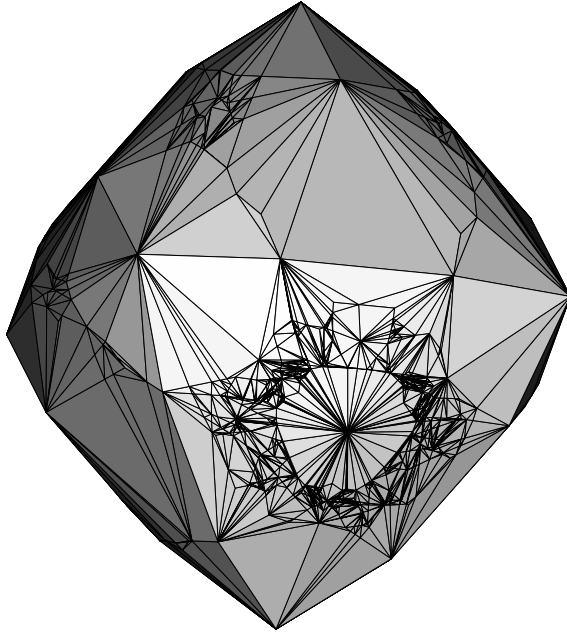
Figure 8: Unit ball in the scl norm in the subspace spanned by
$aabAcAcBCC$, $bbcBaBaCAA$ and $ccaCbCbABB$

of proportionality independent of n. The relative proportion of the image is a constant, so by Theorem 4.1 for any $\epsilon > 0, C > 1$ we have

$$|\operatorname{scl}(v_1 z v_2) \log(n_1 + n_2 + 6)/(n_1 + n_2 + 6) - \log(2k-1)/6| \leq \epsilon$$

with probability $1 - O(n_1^{-C})$. For $n_1 \geq n_2$ large, $\log(n_1 + n_2 + 6)$ is very close to $\log(n_1)$. On the other hand, $|\operatorname{scl}(v_1 + v_2) - \operatorname{scl}(v_1 z v_2)| \leq$ const. It follows for any $\epsilon > 0, C > 1$, with probability $1 - O(n_1^{-C})$,

$$|\operatorname{scl}(v_1 + v_2) - \operatorname{scl}(v_1) - \operatorname{scl}(v_2)| \leq \epsilon n / \log(n).$$

In particular, the boundary of the unit ball contains a point which is very close to the midpoint of the points $v_1/\operatorname{scl}(v_1)$ and $v_2/\operatorname{scl}(v_2)$, and by convexity, the unit ball in the positive quadrant of the v_1, v_2 plane is C^0 close to a triangle. Replacing v_i by v_i^{-1}, the entire unit ball in the v_1, v_2 plane is C^0 close to a diamond. The higher dimensional case is completely analogous. $\square$

5 Quasimorphism lower bound

In this section we exhibit an explicit quasimorphism which certifies a uniform lower bound for scl of a random word. Unfortunately, this lower bound is not sharp, for it exhibits only $\mathrm{scl}(v) \geq n \log(2k-1)/12 \log(n)$ (with high probability), which is $1/2$ of the correct value, by Theorem 4.1.

Experience shows that constructing explicit extremal quasimorphisms is difficult. For example, there is a polynomial time algorithm to produce an extremal surface for a chain in a free group, whereas there is no known algorithm (of any kind) to produce a certifying quasimorphism. Björklund and Hartnick [2] proved a central limit theorem for quasimorphisms (on random walks; but these are very similar to random words in the special case of free groups), and consequently any fixed quasimorphism on F takes values of order $O(\sqrt{n})$ on words of length n. For this reason, it is interesting to be able to construct an explicit quasimorphism that gives the correct $O(n/\log(n))$ order of magnitude. Another nice feature of the construction is that the bound in probability is exponential in n, in contrast to the polynomial bound in Proposition 4.9.

5.1 Quasimorphisms and Bavard Duality

A reference for the material in this section is [5], especially Chapter 2.

Definition 5.1 Let G be a group a *quasimorphism* is a function for which there is a least non-negative real number $D(\phi)$ (called the *defect*) for which

$$|\phi(gh) - \phi(g) - \phi(h)| \leq D(\phi)$$

for all $g, h \in G$.

Furthermore, a quasimorphism is *homogeneous* if $\phi(g^n) = n\phi(g)$ for all $g \in G$ and all integers n.

If ϕ is any quasimorphism, the *homogenization* of ϕ, denoted $\bar{\phi}$, is defined by

$$\bar{\phi}(g) := \lim_{n \to \infty} \phi(g^n)/n.$$

It is a fact that $\bar{\phi}$ is a homogeneous quasimorphism, and satisfies $D(\bar{\phi}) \leq 2D(\phi)$. See [5, Lemma 2.58]. The set of homogeneous quasimorphisms on G is a real vector space $Q(G)$. The subspace with $D = 0$ consists precisely of the homomorphisms $H^1(G; \mathbb{R})$, and D makes the quotient Q/H^1 into a *Banach space*.

There is a duality between quasimorphisms and stable commutator length, known as *Generalized Bavard Duality*. The statement of this duality theorem is:

Theorem 5.2 (Generalized Bavard Duality [5, Theorem 2.79]) *Let G be a group. Then for any $\sum t_i g_i \in B_1^H(G)$ there is an equality*

$$\mathrm{scl}\left(\sum_i t_i g_i\right) = \frac{1}{2} \sup_{\phi \in Q/H^1} \frac{\sum_i t_i \phi(g_i)}{D(\phi)}.$$

A special case of this theorem was established by Bavard in [1]. Notice that this theorem is "complementary" to Lemma 3.4: an admissible surface certifies an upper bound for scl, whereas a homogeneous quasimorphism certifies a lower bound.

An important and useful class of quasimorphisms are the (big) counting quasimorphisms, defined by Rhemtulla [25], and rediscovered by Brooks [3]. Recall the definition of the counting functions C_σ from Section 2.3 and their antisymmetrization $H_\sigma := C_\sigma - C_{\sigma^{-1}}$. Given a set of reduced words $S \subset F$, the function $H_S := \sum_{\sigma \in S} H_\sigma$ is a quasimorphism, and its value on v counts the difference in the number of copies of σ and of σ^{-1} for each $\sigma \in S$. The homogenization counts the difference of the number of copies in the (cyclically reduced) *cyclic* word v.

While big counting quasimorphisms are intuitively very natural, it will be technically easier for us to work with *small* counting quasimorphisms. As above, let $S \subset F$, and define

$$c_S = \text{maximal number of disjoint copies of elements of } S \text{ in } v.$$

Then $h_S = c_S - c_{S^{-1}}$ is a quasimorphism, the small counting quasimorphism on S. See eg [5, Section 2.3.2]. In contrast to big counting quasimorphisms, for which bounding the defect proves difficult, small counting quasimorphisms have a uniformly bounded defect.

Lemma 5.3 *For any $S \subseteq F$, we have $D(h_S) \leq 3$ and $D(\bar{h}_S) \leq 6$.*

Proof This is [9, Lemma 5.1]. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

5.2 Construction of the quasimorphism

Proposition 5.4 *Let v be a random reduced word in the commutator subgroup of length n. Then there is an explicit construction of a homogeneous quasimorphism, so that for all $\epsilon > 0$ there are constants $C > 1$ and $c > 0$ such that with probability $1 - O(C^{-n^c})$, the quasimorphism certifies the inequality*

$$\mathrm{scl}(v) \geq \frac{1}{1+\epsilon} \frac{n \log(2k-1)}{12 \log(n)}.$$

Proof Recall our notation $m = \log(n)/\log(2k-1)$ where k is the rank of the free group F. Fix $L = 1 + \epsilon$ for $\epsilon > 0$, and partition the cyclic word v into adjacent disjoint subwords of length Lm. Note that there may be some small remainder if Lm does not divide n; ignore this gap, as it will be insignificant for our purposes. Let S be the collection of these subwords.

Lemma 5.5 *For $L = 1 + \epsilon$ and S as above, there exist $C > 1$ and $c > 0$ such that with probability $1 - O(C^{-n^c})$, there is a subset $S' \subset S$ with*

$$\mathrm{card}(S - S') < n^{2-L+\epsilon/2}$$

such that for no $\sigma \in S'$ does σ^{-1} appear in v.

Proof Repeating the content of Section 2.5 while assuming that the words in S are disjoint only simplifies the arguments, so Proposition 2.6 still holds in this case. □

The certifying quasimorphism will be $\bar{h}_{S'}$. By construction,

$$\bar{h}_{S'}(v) \geq \frac{n}{Lm} - n^{2-L+\epsilon/2} - 1 - c_{(S')^{-1}}(v),$$

and $c_{(S')^{-1}}(v) = 0$ by Lemma 5.5. By Lemma 5.3, $D(\bar{h}_{S'}) \leq 6$, so Bavard duality gives

$$\mathrm{scl}(v) \geq \frac{\bar{h}_{S'}(v)}{2D(\bar{h}_{S'})} \geq \frac{1}{1+\epsilon} \frac{n \log(2k-1)}{12 \log(n)} - o(n/\log(n)).$$

The statement of the lemma is obtained by repeating the argument with $\epsilon/2$; the multiplicative factor $1/(1+\epsilon)$ then renders the $o(n/\log(n))$ unnecessary. □

6 Computer experiments and a surprisingly good heuristic

Recall that in the proof of Proposition 4.2 we constructed a surface by gluing random tripods. The length of the edges of the tripods was $m/2 = \log(n)/2\log(2k-1)$, but each edge of each tripod was extended to a maximal joint before gluing. If $u = xy$ and $u' = xy'$ are reduced words with a common nonempty prefix x, the expected length of the common prefix of y is $1/(2k-1) + 1/(2k-1)^2 + \cdots = 1/(2k-2)$. This suggests that the average edge length of an extremal surface should be at least $m/2 + 1/(2k-2)$, and therefore that the value of scl should be at most $n/12(\log(n)/2\log(2k-1) + 1/(2k-2))^{-1}$.

Without a really sound theoretical justification, we nevertheless made the prediction that this heuristic correction should more accurately match the actual average value of scl, and tested this experimentally.

Figure 9 displays the result of computer experiment. We computed the scl of 20 random words in $[F_2, F_2]$ of lengths between 70 and 240 (inclusive) in steps of 10. The upper solid line indicates the theoretical value $n \log(2k-1)/6 \log(n)$ from Theorem 4.1, the dots are the actual averages, and the lower dashed line (passing in a very satisfying way through the experimental dots!) is the heuristic $n/12(\log(n)/2 \log(2k-1) + 1/(2k-2))^{-1}$.
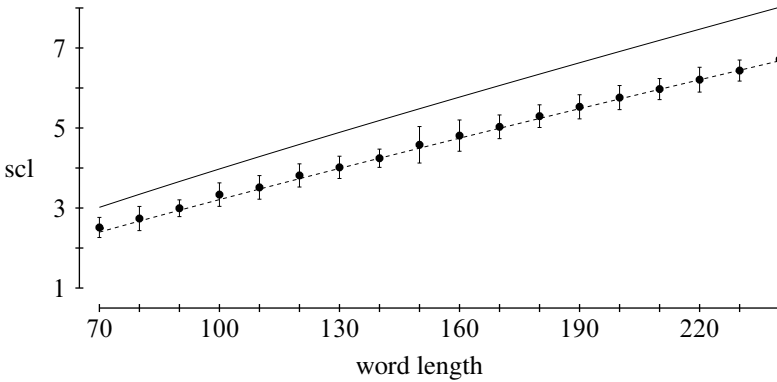

Figure 9: Experimental computation of scl on random words in F_2 of length between 70 and 240, and comparison with (asymptotic) theoretical and heuristic values.

Appendix A: Directed graphs and Markov chains

The purpose of this appendix is firstly to put the estimates obtained in Section 2 into the more general context of the theory of nonreversible Markov chains, and secondly to indicate which aspects of the theory developed above can be expected to generalize easily to hyperbolic groups and spaces, and which aspects require new ideas. The main results of the paper do not depend logically on the results or conjectures in this appendix.

Let X_1 be the directed graph whose vertices are the generators of F, and whose (directed) edges are the (ordered) non-inverse pairs. A random word v of length n can be interpreted as a random walk on X (where edges have the uniform probability) starting at a random vertex (also with the uniform probability). This graph is *ergodic* (ie, there is a directed path from any vertex to any other vertex) and *aperiodic* (ie, the gcd of the lengths of the directed loops is 1).

For any i let X_i be the directed graph whose vertices are the elements of F_i, and whose (directed) edges are the elements of F_{i+1}, where an edge g starts/ends at its

prefix/suffix respectively of length i. Note for each i that X_i is $(2k-1)$–regular, ergodic and aperiodic. Again, a random word v of length n can be interpreted as a random walk on X of length $n-i+1$ starting at a random vertex.

Each X_i determines a nonreversible Markov chain (in the obvious way), with stationary probability π the uniform probability measure on vertices (ie, such that each vertex has weight $1/(2k)(2k-1)^{i-1}$), and Markov kernel $P_i(x,y) = 1/(2k-1)$ if there is a directed edge from x to y; ie, if x and y are reduced words of length i, and the suffix of x of length $i-1$ is equal to the prefix of y of length $i-1$.

For an introduction to the theory of Markov chains, see Freedman [16]. We remark that we use only the most elementary aspects of the theory in this paper, since our Markov chains always have discrete time and finite state space.

A.1: Chernoff inequalities for nonreversible Markov chains

We would like to estimate the rate of convergence of random sums to the equilibrium; that is, we want to estimate the probability that $|n^{-1}\sum_{j=1}^n f(x_j) - \int f\,d\pi|$ is bigger than $n^{-\delta}$, for some function f on the vertices of X_i (ie, on F_i). In the sequel we denote $\int f\,d\pi$ by $\mathrm{mean}_\pi(f)$, or just $\mathrm{mean}(f)$ if π is understood.

As is well-known, for *reversible* Markov chains, the rate of convergence is governed by the spectral gap (ie, the difference between 1 and the second largest eigenvalue) of the (symmetric) Markov kernel P. For *nonreversible* Markov chains, the relevant quantity is the smallest nonzero eigenvalue λ_1 of $L := \mathrm{Id} - (P + P^*)/2$. In general P^* is defined by $P^*(x,y) = \pi(y)P(y,x)/\pi(x)$, so in our context P^* is just the transpose P^T.

Let f be normalized to have $\|f - \mathrm{mean}(f)\|_\infty \le 1$ and $\|f - \mathrm{mean}(f)\|_2^2 \le 1$. Let q be an initial distribution, and define $N_q = \|q/\pi\|_2$ (note that we always have $N_q \le \min(\pi(v))^{-1/2}$). Then the main Chernoff-type inequality, due to Lezaud, is as follows:

Theorem A.1 (Lezaud [22, Theorem 1.1 (cf Remark 1.3)]) *With notation as above, there is an inequality:*

$$\mathbf{P}\Big(n^{-1}\sum_{j=1}^n f(x_j) - \mathrm{mean}(f) \ge \gamma\Big) \le N_q e^{-\lambda_1 n\gamma^2/8}$$

Remark A.2 Replacing f by $-f$ gives the same bound on

$$\mathbf{P}\Big(n^{-1}\sum_{j=1}^n f(x_j) - \mathrm{mean}(f) \le -\gamma\Big).$$

Remark A.3 It is possible to control the rate of convergence in terms of other kinds of spectral data, for instance, the second smallest eigenvalue λ_1 of $\mathrm{Id} - PP^*$. However for the Markov chains X_i as above with $i \geq 2$, the multiplicative reversibilization PP^* has many distinct eigenvectors of eigenvalue 1, so $\lambda_1 = 0$. Another approach is to work directly with the smallest positive *singular value* of the (nonsymmetric) matrix $\mathrm{Id} - P$; this approach is favored by Dinwoodie [15].

Remark A.4 Lezaud's estimate is not in itself strong enough to derive Proposition 2.3 because the variance of a counting function C_σ is too big. Nevertheless, our proof of Proposition 2.3 owes something to the approach of Lezaud, and also to the earlier work of Dinwoodie [15] mentioned above (especially the implicit estimate of the random covering time in Lemma 2.4).

A.2: Estimating λ_1

The following estimate on λ_1 in terms of the spectrum of P is obtained by Chung:

Theorem A.5 (Chung [11, Theorem 4.3]) *If X is a directed graph, the eigenvalue λ_1 of L is related to the (ordered) eigenvalues ρ_i of P as follows:*

$$\min_{i \neq 0}(1 - |\rho_i|) \leq \lambda_1 \leq \min_{i \neq 0}(1 - Re(\rho_i)).$$

Remark A.6 Note that Chung proves her theorem for arbitrary (not necessarily regular) graphs, in which case the Laplacian L has the more complicated form

$$L = \mathrm{Id} - \frac{\Phi^{1/2} P \Phi^{-1/2} + \Phi^{-1/2} P^* \Phi^{1/2}}{2},$$

where Φ is the diagonal matrix whose entries are the values of π. For a regular graph, Φ is a scalar multiple of the identity and $P^* = P^T$, so this simplifies to $\mathrm{Id} - (P + P^T)/2$, which agrees with the definition of L in Theorem A.1.

Lemma A.7 *For $L = \mathrm{Id} - (P_i + P_i^T)/2$ where P_i is the probability matrix for X_i, there is an estimate $\lambda_1 \geq const. > 0$ where const. does not depend on i.*

Proof By Theorem A.5, it suffices to obtain upper bounds on the absolute values $|\rho_i|$ of the spectrum of P_i. But the spectrum of P_i is equal to the spectrum of P_1 for any i (padded by zeros), since the traces of all powers P_i^j and P_1^j are equal. To see this, observe that these traces count the number of periodic cycles in X_i and X_1 of period j, but such cycles in either case are in bijection with bi-infinite periodic words with period j.

So it suffices to show that the spectrum of P_1 has a unique eigenvalue 1 and all other eigenvalues strictly less than 1 in absolute value. This follows from the aperiodicity and ergodicity of X_1. □

Incidentally, X_1 *is* a reversible Markov chain, and therefore the spectrum of P_1 is real, so the same is true for the spectrum of all P_i.

A.3: Cheeger constants in X_i

There are other methods to estimate λ_1 for a directed graph, via a generalization of the classical *Cheeger's inequality*. If X is a regular directed graph, the *Cheeger constant* $h(X)$ is the infimum of $|\partial U|/|U|$ over all subsets U of vertices of X with cardinality at most $|U| \leq |X|/2$, where ∂U is the set of elements of the complement U^c joined by a directed edge from U to U^c.

The significance of this quantity for λ_1 is the following theorem of Chung:

Theorem A.8 (Chung [11, Theorem 5.1]) *Let X be a directed graph. Then*

$$2h(X) \geq \lambda_1 \geq h^2(X)/2.$$

For the sake of interest, we show that the Cheeger constants of the X_i are all equal, which gives another proof of Lemma A.7.

Lemma A.9 *For any i, there is an equality $h(X_i) \geq h(X_1)$.*

Proof We give a sketch of a proof.

Given U a subset of X_i with $|U| \leq |X_i|/2$, let V denote the set of suffixes of U of length $i-1$, and let V' denote the set of words obtained from V by appending a letter. Then $\partial U = V' \backslash U$. Also, let $'V$ denote the set of words obtained from V by prepending a letter. Then $|V'| = |'V| = |V|(2k-1)$ and $U \subset 'V$. Choose U so that

$$|\partial U| = |V' \backslash U| = h(X_i)|U| \leq h(X_i)|V'|.$$

Note that either $|V| \leq |X_{i-1}|/2$, or else we may obtain a lower bound on $h(X_i)$ from the difference $|V| - |X_{i-1}|/2$; for the sake of argument, therefore, assume the former.

Now think of V as a subset of X_{i-1}, and let W denote the set of suffixes of V of length $i-2$, and define W' and $'W$ analogously to above. Then $\partial V = W' \backslash V$ by definition. Moreover, $|W' \backslash V|(2k-1) = |V' \backslash 'V|$ since each element of $W' \backslash V$ can be prepended with $(2k-1)$ different letters to produce an element of $V' \backslash 'V$. Since also $|V|(2k-1) = |V'| \geq |U|$ we deduce

$$h(X_{i-1}) \leq \frac{|\partial V|}{|V|} = \frac{|W' \backslash V|}{|V|} = \frac{|V' \backslash 'V|}{|V'|} \leq \frac{|V' \backslash 'V|}{|U|} \leq \frac{|V' \backslash U|}{|U|} = h(X_i).$$
□

A.4: Hyperbolic groups

For an introduction to hyperbolic groups, see Gromov [18]. A finitely generated group G is *hyperbolic* if it is coarsely negatively curved on a large scale. This can be expressed in several equivalent ways in terms of the geometry of the Cayley graph; the most useful characterizations are

(1) δ–thinness of triangles,

(2) a linear isoperimetric inequality, and

(3) all asymptotic cones are $\mathbb{R}$–trees.

The adjective "hyperbolic" comes from the close (metric) resemblance to hyperbolic geometry. But there is another sense in which such groups are hyperbolic, namely in the dynamics of the (symbolic) geodesic flow.

Cannon [10] showed that in hyperbolic groups, a set of representative shortest words in any given generating set can be enumerated by a finite state automaton. In the language of digraphs, one version of Cannon's theorem can be expressed as follows.

Let G be a hyperbolic group with a symmetric generating set S. Let Γ be a finite directed graph with a distinguished (initial) vertex, and edges labeled by elements of S, in such a way that there is at most one edge with a given label emanating from each vertex. A directed path γ in Γ starting at the initial vertex determines a word $w(\gamma)$ in the generators S, and by evaluation, an element of G. Cannon shows that one can find such a Γ for which there is a 1–1 correspondence between such directed paths and elements of G, and moreover for which every word $w(\gamma)$ is a geodesic, ie it is of shortest length among all words in S^* representing a given element of G. In more geometric terms, let $\widetilde{\Gamma}$ denote the universal cover of Γ (it is also a directed graph), and let Γ' be the subgraph of $\widetilde{\Gamma}$ that is the union of all directed rays starting at some lift of the initial vertex. Then Γ' embeds in the Cayley graph $C_S(G)$ in an edge-label-respecting way as a spanning tree, and every directed path in Γ' is a geodesic in $C_S(G)$.

In this language, there is a correspondence between "random" words in G, and "random" directed walks in Γ. One thinks of Γ as a topological Markov chain, and then one can assign probabilities to the edges (the transitions between states) in a way that maximizes the entropy. For such an assignment, the pushforward measure from walks of length n to the sphere of radius n in $C_S(G)$ is coarsely equivalent to the uniform measure on the sphere, and the limit as $n \to \infty$ converges to the Patterson–Sullivan measure on the Gromov boundary ∂G (see eg Coornaert and Papadopoulos [13]).

A significant technical issue is that the graph Γ is not typically *ergodic*. Given a general directed graph Γ, one can form a new directed graph without cycles, whose vertices are the "communicating classes" of vertices in Γ (ie, equivalence classes of the relation $\sim$, where $u \sim v$ if there is a directed path from u to v and another directed path from v to u). Each vertex of the new graph corresponds to an ergodic subgraph of Γ, whose adjacency matrix has a real, non-negative (Perron–Frobenius) eigenvalue.

From the point of view of probability theory, only the vertices with maximal eigenvalue are significant. It is an important consequence of a theorem of Coornaert [12] that for hyperbolic groups, such vertices do not occur in series, but only in parallel. It follows that this maximal eigenvalue λ is also the growth rate of the group; ie, the unique λ such that there are $\Theta(\lambda^n)$ words of length n. The fact that such "maximal" vertices only occur in parallel means informally that there are finitely many distinct classes W so that all but $O(C^{-n^c})$ words of length n fall into one of the classes of W, and for words v in a given class W_i, for each σ of length $L \log(n)/\log(\lambda)$ with $L < 1$, there is some $f_i(\sigma)$ (depending only on σ and on the class W_i) so that

$$\mathbf{P}\big(|C_\sigma(v) - n f_i(\sigma)| < n^{\epsilon+(1-L)/2}\big) = 1 - O(C^{-n^c}),$$

ie, the analogue of Proposition 2.3 holds for each class W_i separately, and with essentially the same proof. This leaves two problems before one can attempt to generalize the construction in Section 4 to arbitrary hyperbolic groups: one must be able to compare $f_i(\sigma)$ for different classes i, and one must be able to compare $f_i(\sigma)$ with $f_i(\sigma^{-1})$. These problems are largely solved by the methods of Calegari and Fujiwara [7], and Calegari and Maher [8]; see especially [8, Section 3.7].

We believe that it should be straightforward (albeit technically involved) to generalize the results of Section 4 to arbitrary hyperbolic groups, and therefore feel confident in the following conjecture:

Conjecture A.10 Let G be a hyperbolic group with finite generating set S, and let λ be such that the number of elements of length n is $\Theta(\lambda^n)$. Let v be a random element of word length n, conditioned to lie in the commutator subgroup $[G, G]$. Then for any $\epsilon > 0$ and $C > 1$,

$$|\operatorname{scl}(v) \log(n)/n - \log(\lambda)/6| \le \epsilon$$

with probability $1 - O(n^{-C})$.

A similar analogue of Theorem 4.16 should also hold.

A.5: Hyperbolic manifolds

If M is a closed hyperbolic d–manifold, it makes sense to study the stable commutator length of random closed geodesics with length in $[n-\delta, n+\delta]$ for some fixed δ (conditioned to be homologically trivial). The geodesic flow on a hyperbolic manifold is the canonical example of an Anosov flow, and the analogues of Lezaud's Chernoff-type inequality are the mixing theorems of Pollicott [24] and others.

The correct analogue of $\log(\lambda)$ should be the exponential growth rate of the number of orbits as a function of length which is just $d - 1$ (ie, the volume entropy) where d is the dimension. The following conjecture seems very reasonable:

Conjecture A.11 Let M be a closed hyperbolic d–manifold. Fix some $\delta > 0$. Let γ be a random geodesic of length in $[n-\delta, n+\delta]$ conditioned to be homologically trivial, and let v be the corresponding conjugacy class in $\pi_1(M)$. Then for any $\epsilon > 0$ and $C > 1$,
$$|\operatorname{scl}(v)\log(n)/n - (d-1)/6| \leq \epsilon$$
with probability $1 - O(n^{-C})$.

If true, this conjecture would say that one can recover (to any desired accuracy) the *length* of a random geodesic directly from the bounded cohomology of $\pi_1(M)$; this interpretation is obviously very close to the spirit of Gromov's celebrated result discussed in the introduction.

References

[1] **C Bavard**, *Longueur stable des commutateurs*, Enseign. Math. 37 (1991) 109–150 MR1115747

[2] **M Björklund**, **T Hartnick**, *Biharmonic functions on groups and limit theorems for quasimorphisms along random walks*, Geom. Topol. 15 (2011) 123–143 MR2764114

[3] **R Brooks**, *Some remarks on bounded cohomology*, from: "Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference", (I Kra, B Maskit, editors), Ann. of Math. Stud. 97, Princeton Univ. Press (1981) 53–63 MR624804

[4] **M Bucher-Karlsson**, *The simplicial volume of closed manifolds covered by $\mathbb{H}^2 \times \mathbb{H}^2$*, J. Topol. 1 (2008) 584–602 MR2417444

[5] **D Calegari**, *scl*, MSJ Memoirs 20, Mathematical Society of Japan, Tokyo (2009) MR2527432

[6] **D Calegari**, *Stable commutator length is rational in free groups*, J. Amer. Math. Soc. 22 (2009) 941–961 MR2525776

[7]   **D Calegari**, **K Fujiwara**, *Stable commutator length in word-hyperbolic groups*, Groups Geom. Dyn. 4 (2010) 59–90   MR2566301

[8]   **D Calegari**, **J Maher**, *Statistics and compression of scl*, to appear in Ergodic Theory Dyn. Systems

[9]   **D Calegari**, **A Walker**, *Isometric endomorphisms of free groups*, New York J. Math. 17 (2011) 713–743   MR2851070

[10] **J W Cannon**, *The combinatorial structure of cocompact discrete hyperbolic groups*, Geom. Dedicata 16 (1984) 123–148   MR758901

[11] **F Chung**, *Laplacians and the Cheeger inequality for directed graphs*, Ann. Comb. 9 (2005) 1–19   MR2135772

[12] **M Coornaert**, *Mesures de Patterson–Sullivan sur le bord d'un espace hyperbolique au sens de Gromov*, Pacific J. Math. 159 (1993) 241–270   MR1214072

[13] **M Coornaert**, **A Papadopoulos**, *Symbolic dynamics and hyperbolic groups*, Lecture Notes in Mathematics 1539, Springer, Berlin (1993)   MR1222644

[14] **M Culler**, *Using surfaces to solve equations in free groups*, Topology 20 (1981) 133–145   MR605653

[15] **I H Dinwoodie**, *Expectations for nonreversible Markov chains*, J. Math. Anal. Appl. 220 (1998) 585–596   MR1614916

[16] **D Freedman**, *Markov chains*, Springer, New York (1983)   MR686269

[17] **M Gromov**, *Hyperbolic manifolds (according to Thurston and Jørgensen)*, from: "Bourbaki Seminar, Vol. 1979/80", Lecture Notes in Math. 842, Springer, Berlin (1981) 40–53   MR636516

[18] **M Gromov**, *Hyperbolic groups*, from: "Essays in group theory", (S M Gersten, editor), Math. Sci. Res. Inst. Publ. 8, Springer, New York (1987) 75–263   MR919829

[19] **U Haagerup**, **H J Munkholm**, *Simplices of maximal volume in hyperbolic n–space*, Acta Math. 147 (1981) 1–11   MR631085

[20] **J Kahn**, **V Markovic**, *Immersing almost geodesic surfaces in a closed hyperbolic three manifold*, Ann. of Math. 175 (2012) 1127–1190   MR2912704

[21] **D Knuth**, *Big Omicron and big Omega and big Theta*, ACM SIGACT 8 (1976) 18–24

[22] **P Lezaud**, *Chernoff and Berry–Esséen inequalities for Markov processes*, ESAIM Probab. Statist. 5 (2001) 183–201   MR1875670

[23] **R C Penner**, *Perturbative series and the moduli space of Riemann surfaces*, J. Differential Geom. 27 (1988) 35–53   MR918455

[24] **M Pollicott**, *Exponential mixing for the geodesic flow on hyperbolic three-manifolds*, J. Statist. Phys. 67 (1992) 667–673   MR1171148

[25] **A H Rhemtulla**, *A problem of bounded expressibility in free products*, Proc. Cambridge Philos. Soc. 64 (1968) 573–584   MR0225889

[26] **R Sharp**, *Local limit theorems for free groups*, Math. Ann. 321 (2001) 889–904 MR1872533

[27] **D W Stroock**, *Probability theory, an analytic view*, Cambridge Univ. Press (1993) MR1267569

[28] **W P Thurston**, *The geometry and topology of* 3*–manifolds*, lecture notes, Princeton University (1978–1981) Available at http://library.msri.org/books/gt3m

Department of Mathematics, University of Chicago
5734 S University Avenue, Chicago, IL 60637, USA

dannyc@math.uchicago.edu,   akwalker@math.uchicago.edu

http://math.uchicago.edu/~dannyc,   http://math.uchicago.edu/~akwalker