Ergod. Th. & Dynam. Sys. (2010), **30**, 1343–1369 © *2009 Cambridge University Press*
doi:10.1017/S0143385709000662 *Printed in the United Kingdom*

Combable functions, quasimorphisms, and the central limit theorem

DANNY CALEGARI† and KOJI FUJIWARA‡

† *Department of Mathematics, Caltech, Pasadena CA, 91125, USA*
(e-mail: dannyc@its.caltech.edu)

‡ *Graduate School of Information Science, Tohoku University, Sendai, Japan*
(e-mail: fujiwara@math.is.tohoku.ac.jp)

(*Received* 17 *May* 2009 *and accepted in revised form* 23 *June* 2009)

Abstract. A function on a discrete group is weakly combable if its discrete derivative with respect to a combing can be calculated by a finite-state automaton. A weakly combable function is bicombable if it is Lipschitz in both the left- and right-invariant word metrics. Examples of bicombable functions on word-hyperbolic groups include:

(1) homomorphisms to $\mathbb{Z}$;
(2) word length with respect to a finite generating set;
(3) most known explicit constructions of quasimorphisms (e.g. the Epstein–Fujiwara counting quasimorphisms).

We show that bicombable functions on word-hyperbolic groups satisfy a *central limit theorem*: if $\overline{\phi}_n$ is the value of ϕ on a random element of word length n (in a certain sense), there are E and σ for which there is convergence in the sense of distribution $n^{-1/2}(\overline{\phi}_n - nE) \to N(0, \sigma)$, where $N(0, \sigma)$ denotes the normal distribution with standard deviation σ. As a corollary, we show that if S_1 and S_2 are any two finite generating sets for G, there is an algebraic number $\lambda_{1,2}$ depending on S_1 and S_2 such that almost every word of length n in the S_1 metric has word length $n \cdot \lambda_{1,2}$ in the S_2 metric, with error of size $O(\sqrt{n})$.

1. *Introduction*

This paper concerns the statistical distribution of values of certain functions on hyperbolic groups, and lies at the intersection of computer science, ergodic theory, and geometry. Since Dehn (cf. [**9**]), or more recently Gromov (e.g. [**13**]), it has been standard practice to study finitely generated groups geometrically, via large-scale properties of their Cayley graphs. Geometric spaces are probed effectively by Lipschitz functions; in a finitely generated group one has the option of probing the geometry of the Cayley graph by functions which are Lipschitz with respect to both the left- and right-invariant metrics simultaneously. One natural source of such functions are *homomorphisms* to $\mathbb{R}$. For instance, Rivin [**22**] and Sharp [**25**] have studied the abelianization map to H_1 on a free group, and derived strong results (see also Picaud [**19**]). Unfortunately, many groups (even

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

hyperbolic groups) admit no non-trivial homomorphisms to $\mathbb{R}$ or $\mathbb{Z}$; however, in many cases such groups admit a rich (uncountable dimensional) family of *quasimorphisms*: informally, homomorphisms up to bounded error. Quasimorphisms arise in the theory of bounded cohomology, and in the context of a number of disparate extremal problems in topology; see e.g. [**4**] for an introduction. The defining property of a quasimorphism implies that it is Lipschitz in both the left- and right-invariant metrics on a group, and therefore one expects to learn a lot about the geometry of a group by studying its quasimorphisms.

A class of groups for which this study is particularly fruitful is that of *word-hyperbolic groups*. As Cannon [**5**] and Gromov [**13**] pointed out, and Coornaert and Papadopolous elaborated [**6–8**], a hyperbolic group can be parameterized (i.e. *combed*) by walks on a finite directed graph, and the (symbolic) dynamics of the (combinatorial) geodesic flow can be recast in terms of the theory of subshifts of finite type. 'Most' groups (in a suitable sense) are word-hyperbolic, so this gives rise to a rich family of natural examples of dynamical systems with hyperbolic dynamics (in the sense of ergodic theory). Underlying this machinery is the fact that geodesics in a hyperbolic group can be recognized by a particularly simple class of computer programs, namely the kind which can be implemented with a *finite-state automaton*.

Putting these two things together, one is naturally led to study *weakly combable functions*—informally, those whose (discrete) derivative can be calculated by a finite-state automaton—and to specialize the study to the class of *bicombable* functions, namely those which are Lipschitz in both the left- and right-invariant metrics.

Many naturally occurring classes of functions on word-hyperbolic groups are bicombable, including:

(1) homomorphisms to $\mathbb{Z}$;
(2) word length with respect to any finite generating set (not necessarily symmetric);
(3) Epstein–Fujiwara counting functions (see [**11**]).

Bicombability of a function does not depend on a particular choice of combing, or even a generating set. Moreover, the set of all bicombable functions on a given group is a free Abelian group (of infinite rank). Although surely interesting in their own right, such functions are inexorably tied to the theory of quasimorphisms, in at least two ways. First, the best-studied class of quasimorphisms on hyperbolic groups, the so-called Epstein–Fujiwara counting functions (i.e. class (3) above), turn out to be bicombable. Second, given any bicombable function ϕ, its antisymmetrization ψ defined by the formula $\psi(g) = \phi(g) - \phi(g^{-1})$ is a quasimorphism. A particularly simple construction is as follows. Let S be a finite generating set for G (as a semigroup), and S^{-1} the set of elements inverse to S in G. Let $\phi_S : G \to \mathbb{Z}$ compute word length with respect to the generating set S, and $\phi_{S^{-1}} : G \to \mathbb{Z}$ compute word length with respect to the generating set S^{-1}. Then both ϕ_S and $\phi_{S^{-1}}$ are bicombable, and their difference $\psi_S := \phi_S - \phi_{S^{-1}}$ is a bicombable quasimorphism. Typically ψ_S will be highly non-trivial.

The main result in this paper is a *central limit theorem* (Theorem 4.25) for bicombable functions on a word-hyperbolic group. If ϕ is a bicombable function on G, and $\overline{\phi}_n$ denotes the value of ϕ on a random element in G of length n (with respect to some fixed generating set S and with a suitable definition of random to be made precise), there are real numbers E and σ so that there is convergence in the sense of distribution

$$n^{-1/2}(\overline{\phi}_n - nE) \to N(0, \sigma),$$

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

where $N(0, \sigma)$ denotes the normal (Gaussian) distribution with standard deviation σ (the possibility $\sigma = 0$ is allowed, in which case $N(0, \sigma)$ just denotes the Dirac distribution supported at the origin). If ϕ is a quasimorphism, then symmetry forces $E = 0$, so as a corollary one concludes that for any bicombable quasimorphism ϕ on a word-hyperbolic group, and for any constant $\epsilon > 0$, there is an N and a K such that for all $n \geq N$, there is a subset G'_n of the set of all words G_n of length n so that $|G'_n|/|G_n| \geq 1 - \epsilon$, and $|\phi(g)| \leq K\sqrt{n}$ for all $g \in G'_n$.

Another corollary (Corollary 4.27) of independent interest to geometric group theorists is the following. Let S_1 and S_2 be two finite generating sets for a non-elementary word-hyperbolic group G. Then there is an algebraic number $\lambda_{1,2}$ such that for any $\epsilon > 0$ there is a constant K and an N so that if G_n denotes the set of elements of length $n \geq N$ in the S_1 metric, there is a subset G'_n with $|G'_n|/|G_n| \geq 1 - \epsilon$, so that for all $g \in G'_n$ there is an inequality

$$||g|_{S_1} - \lambda_{1,2}|g|_{S_2}| \leq K \cdot \sqrt{n}$$

(of course, $|g|_{S_1} = n$ in this formula).

This observation can probably be generalized considerably. For instance, if H is a quasiconvex subgroup of a hyperbolic group G, then word length in G (in some generating set) is bicombable as a function on H. It follows that for any finite generating sets S_H for H and S_G for G, there is some constant λ such that almost every word of length n in H in the S_H metric has length $\lambda \cdot n$ in G in the S_G metric, with error of order $\sqrt{n}$.

We make a remark about the method of proof. In general terms, to derive a central limit theorem for random sums in a dynamical system one needs two ingredients: independence (i.e. hyperbolicity) and recurrence. Informally, one obtains a Gaussian distribution by adding up *independent* samples from the *same* distribution. In our context hyperbolicity, or independence of trials, comes from the finite size of the graph (and the finite memory of the automaton) parameterizing elements of the group. As is well known to the experts, the difficult point is recurrence: a graph parameterizing a combing of a hyperbolic group may typically fail to be recurrent. In fact, it may have many recurrent subgraphs whose adjacency matrices have largest real eigenvalue equal to the largest real eigenvalue for the adjacency matrix of the graph as a whole. In place of recurrence in this graph, we use recurrence *at infinity*, using Coornaert's theorem that the action of G on ∂G is ergodic with respect to a Patterson–Sullivan measure.

Statistical theorems for values of (certain) quasimorphisms on certain groups have been derived by other authors. Horsham–Sharp [**14**] showed recently that a *Hölder continuous* quasimorphism (see §5) on a free group F satisfies a central limit theorem. In quite a different direction, Sarnak [**24**] showed that the values of the *Rademacher function* on $\mathrm{PSL}(2, \mathbb{Z})$, sorted by geodesic length, form a Cauchy distribution. Here the Rademacher function is a correction term to the logarithm of the Dedekind eta function, and arises in elementary number theory. Thinking of $\mathrm{PSL}(2, \mathbb{Z})$ as a subgroup of the group of isometries of the hyperbolic plane, every element whose trace has absolute value greater than two is represented by an isometry with an axis, with a well-defined (geodesic) translation length. The reason for the appearance of the Cauchy distribution instead of a Gaussian distribution has to do with the nonlinear relation between geodesic length and word length in $\mathrm{PSL}(2, \mathbb{Z})$, and the non-compactness of the quotient $\mathbb{H}^2/\mathrm{PSL}(2, \mathbb{Z})$. Of course, the Rademacher function is bicombable, and therefore our main theorem shows that its values

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

have a Gaussian distribution with respect to any word metric on the word-hyperbolic group $\mathrm{PSL}(2, \mathbb{Z})$.

The plan of the paper is as follows. In §2 we introduce basic definitions and the properties of word-hyperbolic groups and finite directed graphs (which we call digraphs, following the computer scientists' convention). In §3 we define combable functions and establish some of their basic properties, proving that the Epstein–Fujiwara counting quasimorphisms on hyperbolic groups are bicombable. In §4 we prove the central limit theorem and derive the main corollaries of interest. Finally, in §5 we briefly discuss the results of Horsham–Sharp.

2. *Groups and digraphs*

2.1. *Hyperbolic groups.*

Definition 2.1. 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 generating set S is *δ-hyperbolic* if the Cayley graph $C_S(G)$ is δ-hyperbolic as a path metric space.

We say that G is *word-hyperbolic* (or simply *hyperbolic*) if there is a $\delta \geq 0$ and a finite generating set S for which $C_S(G)$ is δ-hyperbolic.

A word-hyperbolic group is *elementary* if it contains a finite index cyclic subgroup. If G is non-elementary, it contains many quasi-isometrically embedded non-Abelian free groups of arbitrary rank, by the ping-pong lemma.

We assume that 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 in the following.

THEOREM 2.2. (Gromov; basic properties of δ-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 with X, whose points consist of quasigeodesic rays up to the equivalence relation of being a finite Hausdorff distance apart, which is metrizable. If X is proper, ∂X is compact, and any quasi-isometric embedding $X \to Y$ between hyperbolic spaces induces a continuous map $\partial X \to \partial Y$.*

For proofs and a more substantial discussion, see [**13**].

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

2.2. *Regular languages.* We give a brief overview of the theory of regular languages and automata. There are many conflicting but equivalent definitions in the literature; we use a parsimonious formulation which is adequate for this paper. For a substantial introduction to the theory of regular languages, see [**10**].

Definition 2.3. An *alphabet* is a finite set whose elements are called *letters*. A *word* (or sometimes *string*) on an alphabet is a finite ordered sequence of letters. A *language* is a subset of the set of all words on a fixed alphabet. A language is *prefix closed* if every prefix of an element of the language is also in the language.

If S is an alphabet, the set of all words on S is denoted by S^*, and a language is a subset of S^*.

Convention 2.4. Throughout this paper, if w denotes a word in S^*, then w_i will denote the prefix of length i and $w(i)$ will denote the ith letter. The length of w is denoted by $|w|$. Hence (for example), w_0 denotes the empty word, and $|w_i| = i$.

Definition 2.5. A *finite-state automaton* (or just automaton for short) on a fixed alphabet S is a finite directed graph Γ with a distinguished initial vertex v_1 (the *input state*), and a choice of a distinguished subset Y of the vertices of Γ (the *accept states*), whose oriented edges are labeled by letters of S in such a way that at each vertex there is at most one outgoing edge with any given label.

The vertices are called the *states* of the automaton. A word w in the alphabet determines a directed path γ in an automaton, defined as follows. The path γ starts at the initial vertex at time 0, and proceeds by reading the letters of w one at a time, and moving along the directed edge with the corresponding label if one exists, and halts if not. If the automaton reads to the end of w without halting, the last vertex of γ is the *final state*, and the word is *accepted* if the final state of γ is an accept state, and *rejected* otherwise. In this paper, by convention, if an automaton halts before reaching the end of a word, the word is *rejected*.

Definition 2.6. A language $L \subset S^*$ is *regular* if it is exactly the set of words accepted by some finite-state automaton.

A regular language L is prefix-closed if and only if it is accepted by some finite-state automaton for which every vertex is an accept state. In this paper, we are only concerned with prefix-closed regular languages. Hence, a word w in S^* is accepted if and only if it corresponds to a directed path of length $|w|$.

Convention 2.7. Throughout this paper, if γ denotes a directed path in a graph, then γ_i will denote the initial subpath of length i, and $\gamma(i)$ will denote the ith vertex in the path *after the initial vertex*. The length of γ is denoted by $|\gamma|$. Hence (for example), $\gamma(0)$ denotes the initial vertex.

If Γ is the underlying directed graph of the automaton, we say that Γ *parameterizes* L. If L is prefix closed and parameterized by Γ, there is a natural bijection between elements of L and paths in Γ starting at the initial vertex.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

Remark 2.8. Some authors by convention add a terminal 'fail' state to an automaton, and add directed labeled edges to this fail state so that at every vertex there is a full set of outgoing directed labeled edges, one for each element of the alphabet. In this paper, automata will only be fed accepted words, so this convention is irrelevant.

An automaton is like a computer with finite resources running a specific program. The language in which the program is written only allows *bounded loops*, that is, loops of the form 'for $i = 1$ to 10 do', where the parameters of the loop are explicitly spelled out, in order to certify that the computer will not run out of resources on execution of the program.

Describing an automaton in terms of vertices and edges is like describing a program in terms of machine code. It is more transparent, and the result is both more human-readable and more easily checkable, to describe instead (in more prosaic terms) the program that the automaton carries out.

2.3. *Directed graphs and components.* Now and in the following we refer to directed graphs by the computer scientists' term 'digraphs'. A *pointed digraph* is a digraph with a choice of an initial vertex.

Definition 2.9. Let Γ be a digraph. A vertex v is *accessible* from a vertex u if there is a directed path in Γ from u to v. A *component* is a maximal subgraph such that for any two vertices u, v in the subgraph (not necessarily distinct) the state v is accessible from u.

Remark 2.10. A subgraph in which every vertex is accessible from every other is sometimes called *recurrent*, so a component is just a maximal recurrent subgraph. We use both terms interchangeably here.

Different components in a digraph are disjoint. If Γ is a pointed digraph, associated with Γ is a digraph $C(\Gamma)$ whose vertices are components of Γ, together with an edge from one component to the other if there is a directed path in Γ between the corresponding components which does not pass through any other component. Note that it is entirely possible that $C(\Gamma)$ is empty.

We can enhance $C(\Gamma)$ to a pointed digraph as follows. If the initial vertex v_1 of Γ is contained in a component C, the corresponding vertex C of $C(\Gamma)$ becomes the initial vertex. Otherwise add an initial vertex to $C(\Gamma)$, and join it by a path to each component C that can be reached in Γ by a path from the initial vertex which does not pass through any other component.

LEMMA 2.11. *There are no oriented loops in* $C(\Gamma)$.

Proof. If there is a path from C to itself in $C(\Gamma)$ passing through C', then there is a loop in Γ from some vertex in C to itself passing through C'. Hence, every vertex of C' is accessible from every vertex of C and, conversely, in contrast to the fact that C and C' are maximal. □

2.4. *Combings.* Let G be a group and S a finite subset which generates G as a semigroup. There is an evaluation map $S^* \to G$ taking a word in the alphabet S to the corresponding element in G. We denote this map by $w \to \overline{w}$. Note that a word $w \in S^*$ can

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

equally be thought of as a directed path in the Cayley graph $C_S(G)$ from id to $\overline{w}$. Note, further, that if S is not symmetric (i.e. $S \neq S^{-1}$), the Cayley graph $C_S(G)$ should really be thought of as a digraph.

Convention 2.12. By Convention 2.4, $\overline{w_i}$ denotes the element in G corresponding to the prefix w_i of w. Hence (for example), $\overline{w_0} = \mathrm{id}$ for any w. By Convention 2.7, if γ denotes the corresponding path in $C_S(G)$ from id to $\overline{w}$, then (by abuse of notation) we write $\gamma(i) = \overline{w_i}$ as elements of G.

Definition 2.13. Let G be generated as a semigroup by a finite subset S. A *combing of G with respect to S* is a prefix-closed regular language $L \subset S^*$ for which the evaluation map is a bijection $L \to G$ and such that every element of L is a geodesic.

Geometrically, let Γ be a digraph parameterizing L. Let $\tilde{\Gamma}$ be the universal cover of Γ, let $\tilde{v_1} \in \tilde{\Gamma}$ be a lift of the initial vertex, and let $T \subset \tilde{\Gamma}$ be the subgraph consisting of all points accessible from $\tilde{v_1}$. There is a natural map $T \to C_S(G)$ taking $\tilde{v_1}$ to id and directed paths to directed paths. The statement that L is a combing with respect to S corresponds to the geometric statement that the image of T is a spanning subtree of $C_S(G)$, and that directed paths in T are taken to geodesics in $C_S(G)$.

Hyperbolic groups admit combings with respect to any finite generating set. In fact, let G be hyperbolic and let S be a fixed finite generating set. Choose a total order $\prec$ on the elements of S and, by abuse of notation, let $\prec$ denote the induced lexicographic order on S^*. That is, $w \prec w'$ for words w, w' if there are expressions $w = xsy$, $w' = xs'y'$ for $s, s' \in S$ and $x, y, y' \in S^*$ where $s \prec s'$.

Let $L \subset S^*$ denote the language of *lexicographically first geodesics* in G. That is, for each $g \in G$, a word $w \in S^*$ with $\overline{w} = g$ is in L if and only if $|w| = |g|$, and $w \prec w'$ for all other $w' \in S^*$ with these properties. It is clear that L is prefix-closed and bijective with G.

THEOREM 2.14. (Cannon [**5, 10**]) *Let G be a hyperbolic group, and S a finite generating set. The language of lexicographically first geodesics is regular (and, therefore, defines a combing of G with respect to S).*

Remark 2.15. Note that the choice of a digraph Γ parameterizing a combing L is *not* part of the data of a combing. A given regular language is typically parameterized by infinitely many different digraphs.

Warning. Many different definitions of combing exist in the literature. They are often, but not always, assumed to be bijective, and sometimes to be quasigeodesic. See [**21**] for a survey and some discussion.

3. *Combable functions*

3.1. *Left- and right-invariant metrics.* Let G be a group with finite generating set S (as a semigroup). There are two natural metrics on G associated with S: a *left*-invariant metric d_L which is just the path metric in the Cayley graph $C_S(G)$, and a *right*-invariant metric d_R where $d_R(a, b) = d_L(a^{-1}, b^{-1})$. If $|\cdot|$ denotes word length with respect to S, then

$$d_L(a, b) = |a^{-1}b| \quad \text{and} \quad d_R(a, b) = |ab^{-1}|.$$

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

Remark 3.1. If S is not equal to S^{-1}, then this is not strictly speaking a metric, since it is not symmetric in its arguments.

LEMMA 3.2. *A function $f : G \to \mathbb{R}$ is Lipschitz for the d_L metric if and only if there is a constant C so that for all $a \in G$ and $s \in S$ there is an inequality*

$$|f(as) - f(a)| \leq C.$$

Similarly, f is Lipschitz for the d_R metric if

$$|f(sa) - f(a)| \leq C.$$

Proof. These properties follow immediately from the definitions. □

Remark 3.3. Note the curious intertwining of left and right: a function is Lipschitz in the *left*-invariant metric if and only if it changes by a bounded amount under multiplication on the *right* by a generator, and *vice versa*.

3.2. *Combable functions.* Suppose that G is a group with finite generating set S, and let $L \subset S^*$ be a combing. Let Γ be a digraph parameterizing L. There are natural bijections

$$\text{paths in } \Gamma \text{ starting at the initial vertex} \longleftrightarrow L \longleftrightarrow G.$$

If $w \in L$, and γ denotes the corresponding path in Γ, then recall that Convention 2.7 says that $\gamma(i)$ is the vertex in Γ visited after reading the first i letters of w.

Definition 3.4. Let G be a group with finite symmetric generating set S, and let $L \subset S^*$ be a combing. A function $\phi : G \to \mathbb{Z}$ is *weakly combable with respect to S, L* (or *weakly combable* for short) 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 $w \in L$ there is an equality

$$\phi(\overline{w}) = \sum_{i=0}^{|w|} d\phi(\gamma(i))$$

where γ is the directed path in Γ corresponding to w.

The function ϕ is *combable* if it is weakly combable and Lipschitz in the d_L metric, and it is *bicombable* if it is weakly combable and Lipschitz in both the d_L and the d_R metrics.

In words, a function ϕ is weakly combable if its derivative along elements in a combing can be calculated by an automaton.

Example 3.5. Word length (with respect to a fixed finite generating set S, symmetric or not) is bicombable. In fact, if L is any combing for S, and Γ is any digraph parameterizing L, the function $d\phi$ taking the value one on every (non-initial) vertex of Γ satisfies the desired properties.

Example 3.6. Let F be the free group with generators a, b. Let $S = \{a, b, a^{-1}, b^{-1}\}$, and let L be the language of reduced words in the generating set. The function

$$\phi(w) = \begin{cases} |w| & \text{if } w \text{ begins with } a \\ 0 & \text{otherwise} \end{cases}$$

is combable but not bicombable.

The definition of weakly combable depends on a generating set and a combing.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

Example 3.7. Let $G = \mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ where $\mathbb{Z}$ is generated by a and $\mathbb{Z}/2\mathbb{Z}$ is generated by b. Let $S = a, a^{-1}, b$ and let L be the combing consisting of words of the form a^n and ba^n for different n. Let $S' = b, (ab), (ab)^{-1}$ and let L' be the combing consisting of words of the form $(ab)^n$ and $b(ab)^n$ (each bracketed term denotes a single 'letter' of S'). Let $\phi : G \to \mathbb{Z}$ be the function defined by $\phi(a^n) = n$ and $\phi(ba^n) = 0$. Then ϕ is weakly combable with respect to L, but not with respect to L' since the derivative $d\phi$ is unbounded along words of L'.

However, the property of being (bi-)combable does *not* depend on a generating set, or a choice of combing.

LEMMA 3.8. *Let G be a word-hyperbolic group. Suppose that S, S' are two finite symmetric generating sets, and L, L' bijective geodesic combings in S^* and $(S')^*$, respectively. Then the sets of functions which are (bi-)combable with respect to L and with respect to L' are equal.*

Proof. The property of being Lipschitz in either the left or right metric is independent of a choice of generating set (and does not depend on a combing), so it suffices to prove the lemma for combable functions. Let ϕ be combable with respect to S, L. We show that it is combable with respect to S', L'.

Let w' be a word in L', and let w be the word in L with $\overline{w}' = \overline{w}$. Let γ be a path in $C_S(G)$ corresponding to w, and let γ' be a path in $C_{S'}(G)$ corresponding to w'. Although γ does not make sense as a *path* in $C_{S'}(G)$, nevertheless each vertex $\gamma(i)$ is an element of G, and therefore corresponds to a unique vertex in $C_{S'}(G)$. The 'discrete path' determined by the vertices $\gamma(i)$ is a quasigeodesic in $C_{S'}(G)$, and by the Morse lemma, fellow travels the path γ'. In other words:

(1) there is a constant C_1 so that every $\gamma(i)$ is within distance C_1 of some $\gamma'(j)$, and every $\gamma'(j)$ is within distance C_1 of some $\gamma(i)$;
(2) there is a constant C_2 so that if $\gamma(i)$ is within distance C_1 of $\gamma'(j)$, then $\gamma(k)$ is within distance C_1 of $\gamma'(j+1)$ for some k with $|k - i| \leq C_2$.

Here distances are all measured in $C_{S'}(G)$.

Suppose that we are given some $v' \in L'$. A word v in L is *good* (for v') if it satisfies the following properties. Let α' be the path in $C_{S'}(G)$ associated with v', and α the path in $C_S(G)$ associated with v, and think of the vertices $\alpha(i)$ as vertices in $C_{S'}(G)$. Then v is good if it satisfies properties (1) and (2) above (with α and α' in place of γ and γ'), and if furthermore the distance from $\overline{v}'$ to $\overline{v}$ is at most C_1. Note that if w'_j is a prefix of w', then there is some prefix w_i of w as above that is good.

Let Γ' be a digraph parameterizing L', and let Γ be a digraph parameterizing L for which there is $d\phi$ from the states of Γ to $\mathbb{Z}$ as in Definition 3.4. Let B denote the (set of vertices in the) ball of radius C_1 about id in $C_{S'}(G)$. For each i, let μ_i be the function whose domain is B, and for each $b \in B$ define $\mu_i(b)$ as follows:

(1) if there is no good word $v \in L$ (for w'_i) with $\overline{v} = \overline{w}'_i b$, set $\mu_i(b)$ equal to an 'out-of-range' symbol E; otherwise:
(2) set $\mu_i(b)$ equal to the tuple whose first entry is the value $\phi(\overline{w}'_i b) - \phi(\overline{w}'_i)$ and whose second entry is the state of Γ after reading the word v.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

Note that given b, the word v if it exists is unique, since $L \to G$ is a bijection. Moreover, since ϕ is combable, there is a constant C_3 so that $\phi(\overline{w}'_i b) - \phi(\overline{w}'_i) \in [-C_3, C_3]$. In other words, the range of μ_i is the *finite* set $\mathbb{E} \cup [-C_3, C_3] \times \Gamma$ where for the sake of brevity, and by abuse of notation, we have denoted the set of states of Γ by Γ (this is the only point in the argument where combability, as distinct from weak combability, is used). Finally, observe that $\mu_i(\mathrm{id})$ is never equal to $\mathbb{E}$, since there is a good v for w'_i with $\overline{v} = \overline{w}'_i$, for every i.

We claim that the function μ_i and the letter $w'(i+1)$ together determine the function μ_{i+1}. For each $b \in B$, the function μ_i tells us whether or not there is a good word v with $\overline{v} = \overline{w}'_i b$, and if there is, the state of Γ after reading such a word. There are finitely many words x of length at most C_2 in S^* which can be concatenated to v to obtain a word $vx \in L$, and the set of such x depends only on the state of Γ after reading v. Each suffix x determines a path β in $C_S(G)$ and a discrete path $\overline{v}\beta(j)$ in $C_{S'}(G)$. Whether the composition vx is good for w'_{i+1} or not depends only on the suffix x, so the set of such suffixes can be determined for each $b \in B$, and the set of endpoints determines a subset of vertices $b' \in B$ for which $\mu_{i+1}(b') \neq \mathbb{E}$. For each such b', the state of Γ after reading vx can be determined, and the difference $\phi(\overline{vx}) - \phi(\overline{v})$ can be determined by the combability of ϕ with respect to Γ.

We claim further that the function μ_i and the letter $w'(i+1)$ together determine $\phi(\overline{w}'_{i+1}) - \phi(\overline{w}'_i)$, because there is a unique b and good $v \in L$ with $\overline{v} = \overline{w}'_i b$ which can be extended to $vx \in L$ with $\overline{vx} = \overline{w}'_{i+1}$. As we know the state of Γ after reading v, we know $\phi(\overline{vx}) - \phi(\overline{v})$. Moreover, we know $\phi(\overline{vx}) - \phi(\overline{w}'_i)$, so we know $\phi(\overline{w}'_{i+1}) - \phi(\overline{w}'_i)$ as claimed.

From this data it is straightforward to construct a new Γ'' parameterizing L' so that $d\phi$ exists as a function on (the states of) Γ''. Every word $w' \in L$ and every integer $i \geq 0$ determines a function μ_i, and a state $\gamma'(i) \in \Gamma'$ (by abuse of notation, but in accordance with Convention 2.7, we denote the path in Γ' associated with w' by γ'), and thereby a triple $(\mu_i, \gamma'(i), \phi(\overline{w}'_i) - \phi(\overline{w}'_{i-1}))$ (with the convention $\phi(\overline{w}'_{-1}) = 0$). The vertices of Γ'' are the set of all possible triples that occur, over all $w' \in L'$ and all i; note that this is a *finite* set. The directed edges of Γ'' are the ordered pairs of triples that can occur for some w' and successive indices i, $i+1$. Evidently Γ'' parameterizes L'; moreover, $d\phi$ is the function whose value on a triple as above is the last coordinate. □

The sum or difference of two (bi-)combable functions is (bi-)combable, and the set of all (bi-)combable functions on a fixed G is therefore a countably infinitely generated free Abelian summand of the Abelian group $\mathbb{Z}^G$ of all $\mathbb{Z}$-valued functions on G.

Example 3.9. Word length with respect to a finite generating set S_1 is bicombable with respect to L where L is a combing with respect to another finite generating set S_2 (compare with Example 3.5).

3.3. *Quasimorphisms.*

Definition 3.10. Let G be a group. A *quasimorphism* is a real valued function ϕ on G for which there is a least non-negative real number $D(\phi)$ called the *defect*, such that for

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

all $a, b \in G$ there is an inequality

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

In words, a function is a quasimorphism if it is a homomorphism up to a bounded error.

Quasimorphisms are related to stable commutator length and to two-dimensional bounded cohomology. For a more substantial discussion, see [**1**] or [**4**].

LEMMA 3.11. *A weakly combable quasimorphism is bicombable.*

Proof. For any $g \in G$ and $s \in S$ we have

$$|\phi(gs) - \phi(g)| \leq |\phi(s)| + D(\phi)$$

and

$$|\phi(sg) - \phi(g)| \leq |\phi(s)| + D(\phi).$$

Since s is finite, the conclusion follows. □

Conversely, let ϕ be bicombable. Then we have the following lemma.

LEMMA 3.12. *Let ϕ be bicombable. Then there is a constant C so that if $w \in L$ is expressed as a product of subwords $w = uv$, then $|\phi(\overline{w}) - \phi(\overline{u}) - \phi(\overline{v})| \leq C$.*

Proof. Let $u' \in L$ be any word so that the paths in Γ associated with u and with u' end at the same vertex. Then $\phi(\overline{w}) = \phi(\overline{u}) + \phi(\overline{u'v}) - \phi(\overline{u'})$. Now choose u' of bounded length, and note that the difference of $\phi(\overline{u'v})$ and $\phi(\overline{v})$ is bounded because of the Lipschitz condition. □

Using this lemma we can construct many bicombable quasimorphisms on G.

THEOREM 3.13. *Let G be hyperbolic, and let S be a finite generating set for G as a semigroup. Let ϕ_S be the bicombable function which counts word length with respect to S, and $\phi_{S^{-1}}$ the bicombable function which counts word length with respect to S^{-1}. Then $\psi_S := \phi_S - \phi_{S^{-1}}$ is a bicombable quasimorphism.*

Proof. Since each of ϕ_S and $\phi_{S^{-1}}$ are bicombable, so is their difference. Let g, h be arbitrary. Let $w, u, v \in L$ correspond to gh, g, h, respectively. Then we can write $u = u'x$, $v = yv'$ and $w = w_1 w_2$ as words in L so that $d_L(y, x^{-1}) \leq \delta$, $d_L(u', w_1) \leq \delta$ and $d_R(v', w_n) \leq \delta$ by δ-thinness of triangles. Now apply Lemma 3.12. □

Remark 3.14. Given an arbitrary bicombable function ϕ, one can define ψ by $\psi(g) = \phi(g) - \phi(g^{-1})$. It is not necessarily true that ψ is bicombable, but an argument similar to the proof of Theorem 3.13 shows that ψ is a quasimorphism.

3.4. *Counting quasimorphisms.* Among the earliest known examples of quasimorphisms are counting quasimorphisms on free groups, first introduced by Brooks [**3**]. Brooks' construction was generalized to word-hyperbolic groups by Epstein–Fujiwara [**11**].

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

Definition 3.15. Let G be a hyperbolic group with 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. A *copy* of σ is a translate $a \cdot \sigma$ for some $a \in G$. For α an oriented simplicial path in $C_S(G)$, let $|\alpha|_\sigma$ denote the maximal number of disjoint copies of σ contained in α. For a directed path α in $C_S(G)$ from id to a, define

$$c_\sigma(\alpha) = |a| - (\text{length}(\alpha) - |\alpha|_\sigma)$$

and define

$$c_\sigma(a) = \sup_\alpha c_\sigma(\alpha)$$

where the supremum is taken over all directed paths α in $C_S(G)$ from id to a.

Define a *counting quasimorphism* to be a function of the form

$$\phi_\sigma(a) := c_\sigma(a) - c_{\sigma^{-1}}(a).$$

Remark 3.16. There is no logical necessity to count disjoint copies of σ in α rather than all copies. The latter 'big' counting function agrees with the convention of Brooks whereas the former 'small' counting function agrees with the convention of Epstein–Fujiwara.

In the following, by convention assume that the length of σ is at least two. From the definition, $\phi_\sigma(g) = -\phi_\sigma(g^{-1})$ for all g. A path α which realizes the supremum in the formula above is called a *realizing path* for a. Since the value of each term of the form $|\cdot|$, $\text{length}(\cdot)$, $|\cdot|_\sigma$ is integral, a realizing path exists for every a.

The crucial property of realizing paths for our applications is the following quasigeodesity property.

LEMMA 3.17. (Epstein–Fujiwara [**11**]) *Let σ be a path in $C_S(G)$ of length at least* 2*, and let $a \in G$. A realizing path for a is a K, ϵ-quasigeodesic in $C_S(G)$, where K and ϵ are universal.*

Note that if G is δ-hyperbolic with respect to some S, then a K, ϵ-quasigeodesic is within distance C of an actual geodesic, where C depends on K, ϵ, δ. In other words, a realizing path (asynchronously) fellow travels any geodesic representative of the word.

3.5. *Greedy algorithm.* Let σ be a string. Let k_σ count the maximal number of disjoint copies of σ in a word w, and let k'_σ count disjoint copies of σ in w by using the *greedy algorithm*: that is, there is equality $k'_\sigma(w) = k'_\sigma(v) + 1$ where v is the word obtained from w by deleting the prefix up to and including the first occurrence of σ.

LEMMA 3.18. (Greedy is good) *The functions k_σ and k'_σ are equal.*

Proof. Let w be a shortest word for which $k_\sigma(w)$ and $k'_\sigma(w)$ are not equal. By definition, $k'_\sigma(w) < k_\sigma(w)$ and since w is shortest, $k'_\sigma(w) = k_\sigma(w) - 1$. Since w is shortest, the suffix of w must be a copy of σ which is counted by k_σ but not by k'_σ. Hence, the greedy algorithm must count a copy of σ which overlaps with this copy. So deleting the terminal copy of σ reduces the values of both k_σ and k'_σ by one, in contrast to the hypothesis that w was shortest. □

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

3.6. *Hyperbolic groups.* The purpose of this section is to prove that the Epstein–Fujiwara counting quasimorphisms are bicombable. The proof has a number of structural similarities with the proof of Lemma 3.8. Note that it suffices to show that they are weakly combable, since any quasimorphism is Lipschitz in both the d_L and the d_R metrics.

THEOREM 3.19. *Counting quasimorphisms on hyperbolic groups are bicombable.*

Proof. 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 $L \to G$.

Let σ be a string, and ϕ_σ the associated counting quasimorphism. It suffices to show that c_σ is weakly combable. For the sake of convenience, we abbreviate c_σ to c throughout this proof.

Fix a word $w \in L$, and let γ be the corresponding path in $C_S(G)$. By Lemma 3.17, a realizing path γ' for $\overline{w}$ is a K, ϵ quasigeodesic, and therefore by the Morse lemma, it satisfies the following two conditions (compare with the proof of Lemma 3.8):

(1) there is a constant C_1 so that every $\gamma(i)$ is within distance C_1 of some $\gamma'(j)$, and every $\gamma'(j)$ is within distance C_1 of some $\gamma(i)$;
(2) there is a constant C_2 so that if $\gamma'(i)$ is within distance C_1 of $\gamma(j)$, then $\gamma'(k)$ is within distance C_1 of $\gamma(j+1)$ for some k with $|k-i| \leq C_2$.

Here distances are all measured in $C_S(G)$. Similarly, if α is a geodesic path, say that a K, ϵ quasigeodesic path α' is good for α if they both start at id, if their endpoints are at most distance C_1 apart, and if they satisfy properties (1) and (2) above (with α and α' in place of γ and γ').

Let T denote the set of proper prefixes of σ. Given a (directed) path $\alpha \in C_S(G)$, first find all disjoint copies of σ in α obtained by the greedy algorithm, and let $p(\alpha) \in T$ be the biggest prefix of σ, disjoint from these copies, that is a 'suffix' of (the word corresponding to) α.

Define a function μ_i as follows. Let B denote the (set of vertices in the) ball of radius C_1 in $C_S(G)$ about the identity. For each i, let μ_i be the function whose domain is $B \times T$, and for each $b \in B$ and $t \in T$ define $\mu_i(b, t)$ as follows:

(1) if there is no good path α (for γ_i) with endpoint $\overline{w}_i b$ and $p(\alpha) = t$, set $\mu_i(b, t)$ equal to an 'out of range' symbol E;
(2) otherwise set $\mu_i(b, t)$ equal to tuple whose first term is $|\overline{w}_i b| - |\overline{w}_i|$ (i.e. relative distance to id), and whose second term is the difference $v_i(\overline{w}_i b, t) - c_\sigma(\overline{w}_i)$, where $v_i(\overline{w}_i b, t)$ is the maximum of c_σ on all good paths α as above with $p(\alpha) = t$.

A good path ending at $\overline{w}_i b$ can be concatenated with a short path ending at $\overline{w}_i$ and conversely, so there is a constant C_3 so that μ_i takes values in $\mathrm{E} \cup [-C_1, C_1] \times [-C_3, C_3]$, a finite set. Every good path ending at some $\overline{w}_{i+1} b'$ is obtained by concatenating a good path ending at $\overline{w}_i b$ with a path of length at most C_2, and these can all be enumerated. The relative distance to id on balls of uniform size can be determined as one moves along the geodesic γ, by keeping track of only a bounded amount of data at each stage (this is the heart of Cannon's argument that hyperbolic groups are combable; for details see [**5**]). As in the proof of Lemma 3.8, the function μ_i on $B \times T$ and the letter $w(i+1)$ together determine the function μ_{i+1}, and the difference $c_\sigma(\overline{w}_{i+1}) - c_\sigma(\overline{w}_i)$. From this it

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

readily follows that c_σ is weakly combable, and (since it is evidently bi-Lipschitz) therefore bicombable. The same is true of $c_{-\sigma}$, and therefore of their difference ϕ_σ. □

4. *Ergodic theory of bicombable functions*

Let G be a hyperbolic group with finite generating set S. Let L be a combing with respect to S, and Γ a digraph parameterizing L. Graphs Γ of this kind are not arbitrary, but are in fact quite constrained. The ergodic theory of Γ can be understood using the basic machinery of Markov chains. This material is very well understood, but we keep the discussion self-contained and elementary as far as possible. Basic references for this material include [**2, 12, 16, 20, 23**], and at a couple of points we refer the reader to these texts for proofs of standard facts.

4.1. *Almost semisimple directed graphs.*

Let Γ be a finite directed graph. 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. We can and do think of V as the space of real-valued functions on the vertices of Γ. Let $V \otimes \mathbb{C}$ denote the corresponding vector space of complex-valued functions on the vertices of Γ. Let v_1 denote the initial vertex, and enumerate the other vertices somehow as $v_2, v_3, \ldots$. For a vector $v \in V$, let $|v|$ denote the L^1 norm of v. That is,

$$|v| = \sum_i |\langle v, v_i \rangle|.$$

Definition 4.1. The *transition matrix* of Γ (also sometimes called the *adjacency matrix*) is the $n \times n$ matrix M whose M_{ij} entry counts the number of directed edges from v_i to v_j (or is equal to zero if there are no such edges).

LEMMA 4.2. *For any v_i, v_j the number of directed paths in Γ of length n from v_i to v_j is $\langle v_i, M^n v_j \rangle = (M^n)_{ij}$.*

Proof. This is true for $n = 1$. Suppose, by induction, that it is true for $n = m$. The number of paths of length $m + 1$ from v_i to v_j is equal to the sum over all v_k of the number of paths of length m from v_i to v_k times the number of paths of length one from v_k to v_j. By induction, this is $\sum_k (M^m)_{ik} M_{kj} = (M^{m+1})_{ij}$. □

By Lemma 4.2, the number of paths in Γ of length n starting at v_1 is $|(M^{\mathrm{T}})^n v_1|$, where M^{T} denotes the transpose of M. One could equally well adopt the notation $|v_1 M^n| := |(M^{\mathrm{T}})^n v_1|$ but we prefer the more explicit notation (involving the transpose) since there are many opportunities for left–right errors in what follows.

Definition 4.3. A directed graph 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

$$K^{-1} \lambda^n \leq |(M^{\mathrm{T}})^n v_1| \leq K \lambda^n$$

for all positive integers n.

In what follows we assume that Γ is almost semisimple.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

LEMMA 4.4. *Suppose that* Γ *is almost semisimple. Then* λ *is the largest real eigenvalue of* M. *Moreover, for every eigenvalue* ξ *of* M *either* $|\xi| < \lambda$ *or otherwise 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). We consider $V \otimes \mathbb{C}$ with the hermitian inner product.

Property (2) from Definition 4.3 implies that for any v_i, there is an inequality $|(M^{\mathrm{T}})^n v_i| \leq C_i |(M^{\mathrm{T}})^n v_1|$ 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 $w \in V$ with $|w| = 1$ there is an inequality $|(M^{\mathrm{T}})^n w| \leq C |(M^{\mathrm{T}})^n v_1|$. This is also true if $w \in V \otimes \mathbb{C}$.

For each i, there is some $w_i \in V \otimes \mathbb{C}$ in the ξ_i-eigenspace for which

$$|(M^{\mathrm{T}})^n w_i| \geq \text{constant} \cdot n^{k-1} |\xi_i|^n,$$

where k is the dimension of the Jordan block associated with ξ_i. Since $|(M^{\mathrm{T}})^n w_i| \leq C |(M^{\mathrm{T}})^n v_1|$, by property (3) from Definition 4.3, either $|\xi_i| < \lambda$ or $|\xi_i| = \lambda$ and $k = 1$.

By the Perron–Frobenius theorem for non-negative matrices (see e.g. [**20**, §3.2, pp. 23–26]), M 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 and M^{T} have the same set of eigenvalues, with the same multiplicities. Since every eigenvalue of M with largest (absolute) value has geometric multiplicity equal to its algebraic multiplicity, the same is true of M^{T}.

For any vector $v \in V$, decompose $v = \sum_\xi v(\xi)$ into the components in the generalized eigenspaces of the eigenvalues ξ. Note that $(Mv)(\xi) = M(v(\xi))$, so we write it as $Mv(\xi)$. Since any two norms on $V \otimes \mathbb{C}$ are equivalent, there is a constant $K > 1$ such that

$$K^{-1} \leq \frac{|M^n v|}{\sum_\xi |M^n v(\xi)|} \leq K,$$

providing $M^n v \neq 0$.

LEMMA 4.5. *For any vector* $v \in V$, *the following limit*

$$\rho(v) := \lim_{n \to \infty} n^{-1} \sum_{i=0}^{n} \lambda^{-i} M^i v$$

exists and is equal to $v(\lambda) \in V$.

Proof. We suppress v in the notation that follows. For each eigenvalue ξ define

$$\rho_n(\xi) = n^{-1} \sum_{i=0}^{n} \lambda^{-i} M^i v(\xi)$$

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, and is equal to $v(\lambda)$.

By Lemma 4.4, 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 otherwise the vectors $\lambda^{-i} M^i v(\xi)$ become

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

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 $n^{-1} \sum_{0 \leq i \leq n} \lambda^{-i} M^i v(\xi) \to 0$ unless $\xi = \lambda$, in which case $\rho_n(\lambda) = v(\lambda)$ is constant. □

The same argument as Lemma 4.5 implies.

LEMMA 4.6. *For any vector $v \in V$, the following limit*

$$\ell(v) := \lim_{n \to \infty} n^{-1} \sum_{i=0}^{n} \lambda^{-i} (M^{\mathrm{T}})^i v$$

exists, and is equal to the projection of v 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 and real, so are $\rho(v)$ and $\ell(v)$.

LEMMA 4.7. *For any v, $w \in V$, there is an identity*

$$\langle \ell(v), w \rangle = \langle \ell(v), \rho(w) \rangle = \langle v, \rho(w) \rangle.$$

Proof. Let π denote projection onto the (right) λ-eigenspace, so $\pi v = \rho(v)$ and $\ell(v) = \pi^{\mathrm{T}} v$. Since π is a projection, it is idempotent, i.e. $\pi \circ \pi = \pi$, and similarly for π^{T}. Hence,

$$\langle \pi^{\mathrm{T}} v, w \rangle = \langle \pi^{\mathrm{T}} \pi^{\mathrm{T}} v, w \rangle = \langle \pi^{\mathrm{T}} v, \pi w \rangle = \langle v, \pi \pi w \rangle = \langle v, \pi w \rangle.$$ □

4.2. *Stationary Markov chains.* We would like to be able to discuss the *typical* behavior of an infinite (or sufficiently long) path in Γ. The standard way to do this is to use the machinery of *Markov chains*. A *stationary Markov chain* is a random process $x_0, x_1, x_2, \ldots$ where each x_i is in one of a finite set of states V. The process starts in one state, and successively moves from one state to another. If the chain is in some state $v_i \in V$, then the probability that it moves to state $v_j \in V$ at the next step is some number N_{ij} which depends only on i and j; in other words, given the present state, the future states are independent of the past states. The starting state is determined by some initial probability distribution μ. See [**12**, Ch. 11] for an introduction to the theory of Markov chains.

Our digraph Γ determines a stationary Markov chain as follows. For each n, let X_n denote the set of paths in Γ of length n, starting at an arbitrary vertex, and Y_n the set of paths of length n starting at the initial vertex. A path of length zero is a vertex. Restricting to a prefix defines surjections $X_{n+1} \to X_n$ and $Y_{n+1} \to Y_n$ for all n. Giving each X_n, Y_n the discrete topology, we obtain inverse limits X and Y which denote the space of infinite paths starting at an arbitrary vertex, and infinite paths starting at the initial vertex, respectively.

Topologically, X and Y are Cantor sets. The projections $X \to X_n$ and $Y \to Y_n$ pull back points to open subsets called *cylinders*, which generate the Borel σ-algebras of X and Y.

Definition 4.8. The *shift map* $S : X \to X$ takes an infinite path to its suffix which is the complement of the initial vertex.

It is bothersome to use the same letter S both for the shift map and for a generating set for G, but (unfortunately) both notations are standard, and in any case the meaning should always be clear from context.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

Let $\mathbf{1}$ denote the constant function on Γ taking the value one on every vertex. Define a matrix N by

$$N_{ij} = \frac{M_{ij}\rho(\mathbf{1})_j}{\lambda\rho(\mathbf{1})_i}$$

if $\rho(\mathbf{1})_i \neq 0$ and otherwise define $N_{ii} = 1$ and $N_{ij} = 0$ for $i \neq j$. Define a measure μ on the vertices of Γ by

$$\mu_i = \rho(\mathbf{1})_i \ell(v_1)_i$$

where subscripts denote vector and matrix components.

LEMMA 4.9. *The matrix N is a stochastic matrix (i.e. it is non-negative, and for all i it satisfies $\sum_j N_{ij} = 1$) and preserves the measure μ.*

Proof. For any i not in the support of $\rho(\mathbf{1})$ we have $\sum_j N_{ij} = 1$ by fiat. Otherwise, we have

$$\sum_j N_{ij} = \sum_j \frac{M_{ij}\rho(\mathbf{1})_j}{\lambda\rho(\mathbf{1})_i} = \frac{(M\rho(\mathbf{1}))_i}{\lambda\rho(\mathbf{1})_i} = 1.$$

This shows N is a stochastic matrix. To see that it preserves μ, we calculate

$$\begin{aligned}\sum_i \mu_i N_{ij} &= \sum_i \rho(\mathbf{1})_i \ell(v_1)_i \frac{M_{ij}\rho(\mathbf{1})_j}{\lambda\rho(\mathbf{1})_i} \\ &= \frac{1}{\lambda}\rho(\mathbf{1})_j \sum_i \ell(v_1)_i M_{ij} \\ &= \rho(\mathbf{1})_j \ell(v_1)_j = \mu_j.\end{aligned}$$

Hence, N preserves μ. □

We scale μ to be a probability measure. By abuse of notation, we also denote this probability measure by μ. This probability measure is preserved by N.

There is an associated probability measure on each X_n, which we also denote by μ, where

$$\mu(v_{i_0}v_{i_1}\cdots v_{i_n}) = \mu(v_{i_0})N_{i_0i_1}N_{i_1i_2}\cdots N_{i_{n-1}i_n}.$$

This defines a measure on each cylinder (i.e. the preimage of an element of X_n under $X \to X_n$) and therefore a probability measure μ on X. Lemma 4.9 implies that μ on X is invariant under the shift map S, that is, $\mu(A) = \mu(S^{-1}(A))$ for all measurable $A \subset X$. We call μ the *stationary measure*.

LEMMA 4.10. *Suppose that Γ is almost semisimple, and that μ is the stationary measure on Γ for N. The support of μ is equal to the union of the maximal recurrent components of Γ whose adjacency matrix has the biggest eigenvalue λ. Furthermore, if C, C' are distinct components in the support of μ, there is no directed path from C to C'.*

Proof. Recall that the Landau notation $f(x) = \Theta(g(x))$ for non-negative functions f, g of x means that there are positive constants k_1, k_2 so that

$$k_1 g(x) \leq f(x) \leq k_2 g(x)$$

and the Landau notation $f(x) = o(g(x))$ means that $f(x)/g(x) \to 0$ as $x \to \infty$.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

Since $\mu_i = \rho(\mathbf{1})_i \ell(v_1)_i$, a vertex v_j is in the support of μ if and only if the number of paths from v_1 to v_j of length $\in [n-k, n+k]$ is $\Theta(\lambda^n)$ for some sufficiently big constant k, and the number of outgoing paths from v_j of length $\in [n-k, n+k]$ is also $\Theta(\lambda^n)$. Associated with a recurrent component C of Γ one can form the adjacency matrix of the subgraph C, which has a biggest real eigenvalue $\xi(C)$, by the Perron–Frobenius theorem. Suppose that we have $\xi(C) < \lambda$ for every directed path γ from v_1 to v_j and for every recurrent component C which intersects γ. Then it is easy to see that the number of paths from v_1 to v_j of length $\in [n-k, n+k]$ is $o(\lambda^n)$. Similarly, if for every directed path γ starting at v_j and every recurrent component C' which intersects γ we have $\xi(C') < \lambda$, then the number of outgoing paths from v_j of length $\in [n-k, n+k]$ is $o(\lambda^n)$. Hence, v_j is in the support of μ if and only if there are recurrent components C, C' with $\xi(C) = \xi(C') = \lambda$ such that there is a path from v_1 to v_j intersecting C, and an outgoing path from v_j intersecting C'. Note that $v_j \in C = C'$ is explicitly allowed here. If $C \neq C'$ but there is a path from C to C', then the number of paths of length n which start in C and end in C' has order $\Theta(\sum_i \lambda^i \lambda^{n-i}) = \Theta(n\lambda^n)$, in contrast to the hypothesis that Γ is almost semisimple. It follows necessarily that $C = C'$. □

The construction of the measure μ and the stochastic matrix N might seem unmotivated for the moment. In the next section we provide a more geometric interpretation of these objects.

4.3. *Word-hyperbolic groups and Patterson–Sullivan measures.* We now relate this picture to the theory of hyperbolic groups. Let G be a non-elementary hyperbolic group with generating set S.

Definition 4.11. 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.

THEOREM 4.12. (Coornaert [**6**, Theorem 7.2]) *Let G be a non-elementary 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 4.12 that the critical exponent of the Poincaré series is equal to $\log(\lambda)$, and the series $\zeta_G(\log(\lambda))$ diverges.

Example 4.13. Not every group with a geodesic combing satisfies such an inequality. For example, let $G = F_2 \times F_2$ with the standard generating set. Then we can estimate

$$K^{-1}n3^n \leq |G_n| \leq Kn3^n$$

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

for suitable $K > 0$. On the other hand, the language of words uv where u is a reduced word in the first F_2 factor and v is a reduced word in the second factor defines a geodesic combing.

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 $G \cup \partial G$, where ∂G denotes the Gromov boundary of G.

Definition 4.14. A weak limit $\nu := \lim_{n\to\infty} \nu_n$ is a *Patterson–Sullivan* measure associated with S.

In fact, the limit in Definition 4.14 exists, and an explicit formula for a closely related measure $\widehat{\nu}$ (which differs from ν only by scaling) is given shortly. Note that the support of ν is contained in ∂G, since the Poincaré series diverges at the critical exponent λ.

Let L be a combing of G with respect to S, and let Γ be a digraph. From Coornaert's theorem we immediately deduce the following.

LEMMA 4.15. *Let Γ be as above. Then Γ is almost semisimple.*

A slightly better normalization of the measures ν_n are the sequence of measures $\widehat{\nu}_n$ defined by the formula

$$\widehat{\nu}_n = \frac{1}{n} \sum_{|g|\leq n} \lambda^{-|g|}\delta_g$$

and let $\widehat{\nu}$ be a weak limit (supported on ∂G).

LEMMA 4.16. *Any two weak limits $\nu, \widehat{\nu}$ on ∂G as above are absolutely continuous with respect to each other, and satisfy $1/K \leq d\widehat{\nu}/d\nu \leq K$ for some $K \geq 1$.*

Proof. The lemma is true for each $\widehat{\nu}_n$, ν_n by Theorem 4.12, and therefore it is also true for their limits. □

The measure $\widehat{\nu}$ is not a probability measure in general, although it is finite.

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 4.17. (Coornaert [**6**, Theorem 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 ∂G is ergodic with respect to ν.*

Here we say that a group action which preserves a measure class is *ergodic* if for all subsets $A, B \subset \partial G$ with positive ν-measure, there is $g \in G$ with $\nu(gA \cap B) > 0$. By Lemma 4.16, the action of G on ∂G preserves the measure class of $\widehat{\nu}$ and acts ergodically.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

Remark 4.18. In fact, Coornaert shows that there is some constant $K \geq 1$ such that for any $s \in S$ there is an inequality

$$K^{-1} \leq \frac{d(s_*\nu)}{d\nu} \leq K$$

that is, the action is 'quasiconformal'. It follows that the same is true for the $\widehat{\nu}$ measure, with the same constant K. We do not use this fact in this paper.

We now relate the measure $\widehat{\nu}$ to the measure μ and the stochastic matrix N defined in §4.2.

There is a natural map $Y_n \to G$ for each n, taking a path in Γ to the endpoint of the corresponding geodesic path in G which starts at id. The measures $\widehat{\nu}_n$ determine a measure on Y as follows.

For each $g \in G$, let $\mathrm{cone}(g)$ denote the set of $g' \in G$ such that the geodesic for g' in the combing passes through g. Let $p : Y \to Y_n$ take an infinite path to its prefix of length n. If $y \in Y_n$ corresponds to $g \in G$, define

$$\widehat{\nu}(p^{-1}(y)) = \lim_{n\to\infty} \widehat{\nu}_n(\mathrm{cone}(g)).$$

Since the cylinders $p^{-1}(y)$ define the Borel σ-algebra on Y, this defines a measure $\widehat{\nu}$ on Y. There is a map $Y \to \partial G$ taking an infinite path to the endpoint of the corresponding geodesic in G. The pushforward of $\widehat{\nu}$ on Y under this map agrees with $\widehat{\nu}$ on ∂G.

We give an explicit formula for $\widehat{\nu}_m(\mathrm{cone}(g))$ for any g. By definition, for any $g \in G_n$ and any $m \geq n$ we have

$$\widehat{\nu}_m(\mathrm{cone}(g)) = \frac{1}{m} \sum_{\substack{h \in \mathrm{cone}(g) \\ |h| \leq m}} \lambda^{-|h|}.$$

If $v_g \in \Gamma$ is the vertex corresponding to the endpoint of the path for the element $y \in Y_n$ corresponding to g, then we can rewrite this as

$$\widehat{\nu}_m(\mathrm{cone}(g)) = \frac{1}{m}\lambda^{-n} \sum_{i \leq m-n} \lambda^{-i} \langle (M^{\mathrm{T}})^i v_g, \mathbf{1} \rangle$$

taking a limit as $m \to \infty$, we obtain

$$\widehat{\nu}(p^{-1}(y)) = \lambda^{-n}\langle \ell(v_g), \mathbf{1} \rangle = \lambda^{-n}\langle v_g, \rho(\mathbf{1}) \rangle = \lambda^{-n}\rho(\mathbf{1})_g.$$

Measures are pushed forward by maps, so $S_*\widehat{\nu}$ is the measure on X satisfying

$$S_*\widehat{\nu}(B) = \widehat{\nu}(S^{-1}(B))$$

for any Borel $B \subset X$, and similarly for positive powers S^i.

The subset $Y \subset X$ contains the support of $\widehat{\nu}$. By convention, v_1 has no incoming edges, so Y is disjoint from SY, and therefore $\widehat{\nu}$ is not invariant under S. However, the limit

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{n} S_*^i \widehat{\nu}$$

is manifestly S invariant. The fact that this limit exists is contained in the proof of the next lemma.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

LEMMA 4.19. *The measure* $\lim_{n\to\infty} \sum_{i=1}^{n} (1/n) S_*^i \widehat{\nu}$ *on* X *is equal to the measure* μ *on* X *defined on cylinders using the measure* μ *on* Γ *and* N *as in §4.2.*

Proof. We identify $X_0 = \Gamma$, and observe that the measure μ on Γ agrees with the measure μ on cylinders $p^{-1}(v)$ of X for $v \in \Gamma$. For each vertex v_i of Γ we can calculate

$$S_*^n \widehat{\nu}(p^{-1}(v_i)) = \lambda^{-n} \sum_{\substack{y \in Y_n \\ S^n y = v_i}} \langle v_i, \rho(\mathbf{1}) \rangle.$$

On the other hand, the number of $y \in Y_n$ with $S^n y = v_i$ is exactly equal to the number of directed paths in Γ of length n which start at v_1 and end at v_i, which is $\langle v_1, M^n v_i \rangle$. Hence,

$$\begin{aligned} \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{n} S_*^i \widehat{\nu}(p^{-1}(v_i)) &= \lim_{n\to\infty} \langle v_i, \rho(\mathbf{1}) \rangle \left\langle v_1, \frac{1}{n} \sum_{i \le n} \lambda^{-n} M^n v_i \right\rangle \\ &= \langle v_i, \rho(\mathbf{1}) \rangle \langle v_1, \rho(v_i) \rangle = \rho(\mathbf{1})_i \ell(v_1)_i = \mu_i. \end{aligned}$$

This shows that the measures agree on the cylinders $p^{-1}(v_i)$ for $v_i \in \Gamma = X_0$. By the defining property of μ on X, it suffices to show that for every $y \in Y$ whose nth vertex is v_i, the $\widehat{\nu}$ probability that the $(n+1)$th vertex is v_j is N_{ij}. If y_n is a prefix of y of length n whose last vertex is v_i, and $g \in G_n$ is the corresponding element, we need to calculate

$$\lim_{m\to\infty} \frac{\sum_h \widehat{\nu}_m(\mathrm{cone}(h))}{\widehat{\nu}_m(\mathrm{cone}(g))},$$

where the sum is taken over all $h \in G_{n+1} \cap \mathrm{cone}(g)$ corresponding to $y' \in Y_{n+1}$ whose last vertex is v_j. For each such g, the number of h is M_{ij} so this ratio is equal to $M_{ij}\rho(\mathbf{1})_j / \lambda \rho(\mathbf{1})_i$ which equals N_{ij} as claimed. □

In fact we can be even more precise about the relationship between $\widehat{\nu}$ and μ. If $y \in Y_n$ is a prefix whose last vertex projects to some $v \in \Gamma$ in the support of μ, then S^n maps the cylinder $p^{-1}(y)$ of Y to the cylinder X_v of X consisting of paths with initial vertex v, and from the definitions, the pushforward of $\widehat{\nu}|_{p^{-1}(y)}$ to X under S^n is proportional to $\mu|_{X_v}$.

4.4. *Central limit theorem for stationary Markov chains.* The stationary measure μ on Γ defined in §4.2 has support equal to a finite union of maximal recurrent components of Γ, by Lemma 4.10. Label these components C^i, and let μ^i be the probability measure on C^i which is proportional to $\mu|_{C^i}$. The matrix N preserves μ^i, so it makes sense to talk about a random walk on C^i with respect to N.

Recall that $d\phi$ exists as a function on the vertices of Γ. The stationary Markov chain defined by the matrix N defines a random sequence $x_0, x_1, x_2, \ldots$ of vertices in C^i where x_0 is chosen randomly from the vertices of C^i with respect to the measure μ^i, and at each subsequent stage if x_i corresponds to a vertex v_j, the probability that x_{i+1} will correspond to the vertex v_k is equal to N_{jk}. A sequence $x_0, x_1, x_2, \ldots$ obtained in this way is called a *random walk* on C^i. If X_n^i denotes the subset of X_n consisting of paths of length n which start at some vertex in C^i, then the measure μ on X_n restricted to X_n^i can be scaled to a probability measure μ^i on X_n^i. In this notation, a random walk on C^i of length n corresponds to a random element of X_n^i.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

Let $\overline{S}^i_n$ be the sum of the values of $d\phi$ on a random walk on C^i of length n. Technically, $\overline{S}^i_n$ is a *distribution* on $\mathbb{Z}$. There is a map $X^i_n \to \mathbb{Z}$ taking each walk $x_0, x_1, x_2, \ldots$ to the sum of $d\phi$ on the corresponding vertices in Γ. The pushforward of the measure μ^i on X^i_n is the distribution $\overline{S}^i_n$. In this context, the central limit theorem for stationary (recurrent) Markov chains, due essentially to Markov [**17**], takes the following form.

THEOREM 4.20. (Markov) *Let E^i be the integral of $d\phi$ with respect to the probability measure μ^i. Then there is some $\sigma^i \geq 0$ for which*

$$\lim_{n\to\infty} \mathbf{P}\left(r \leq \frac{\overline{S}^i_n - nE^i}{\sqrt{n(\sigma^i)^2}} \leq s\right) = \frac{1}{\sqrt{2\pi}} \int_r^s e^{-x^2/2}\, dx,$$

where $\mathbf{P}$ *denotes probability with respect to the measure μ^i.*

Stated another way, there is a convergence in the sense of distribution

$$n^{-1/2}(\overline{S}^i_n - nE^i) \to N(0, \sigma^i),$$

where $N(0, \sigma^i)$ denotes a normal (i.e. Gaussian) distribution with mean 0 and standard deviation σ^i. The standard deviation σ^i can be computed easily from μ and N, and is an algebraic function of the entries. Hence, in particular, E^i and σ^i are algebraic and effectively computable.

For a proof, see, for example, [**23**, §46] and see [**23**, §47] for a derivation of a number of formulae for σ^i.

4.5. *Ergodicity at infinity and the central limit theorem for bicombable functions.* To derive a central limit theorem for ϕ, we need to be able to compare means E^i and standard deviations σ^i associated with different components C^i. To do this we introduce the idea of a *typical path* (with respect to the function ϕ).

For any real number r, let $\delta(r)$ denote the probability measure on $\mathbb{R}$ consisting of an atom of mass one concentrated at r. Let γ be an infinite geodesic ray in G. For some fixed real number E and for integers n, m we can define the distribution

$$\omega(n, m)(\gamma) := \sum_{i=1}^{m} \frac{1}{m} \delta((\phi(\gamma_{i+n}) - \phi(\gamma_i) - nE)n^{-1/2}).$$

Definition 4.21. A geodesic ray γ is E, σ-typical if the limit

$$\omega(\gamma) := \lim_{n\to\infty} \lim_{m\to\infty} \omega(n, m)(\gamma)$$

exists in the sense of distribution, and is equal to $N(0, \sigma)$. An element $y \in Y$ is E, σ-typical if the corresponding geodesic ray in G is E, σ-typical.

We parse this condition as follows. For each segment of γ of length n (contained between $\gamma(0)$ and $\gamma(n+m)$), we look at the difference of ϕ on the endpoints. In this way we obtain a set of m numbers, which we think of as a probability measure on $\mathbb{R}$, by weighting each number equally. We then translate the origin by nE, and rescale the x-axis by a factor of $n^{-1/2}$, and consider the resulting probability measure $\omega(n, m)(\gamma)$.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

For fixed n, we consider the limit $\omega(n)(\gamma)$ (if it exists); this distribution is the (rescaled) difference of ϕ on the endpoints of a 'random' segment of γ of length n. Finally, we take the limit as $n \to \infty$ (the rescaling is done so that this limit typically exists). The result $\omega(\gamma)$ is the difference of ϕ on the endpoints of a 'long random' segment of γ, translated and scaled.

For each component C^i, let Y^i be the subset of Y consisting of paths which eventually enter the component C^i and never leave. Note that Y^i and Y^j are disjoint if $i \neq j$.

LEMMA 4.22. *The measure* $\widehat{\nu}(Y - \bigcup_i Y^i) = 0$. *Moreover,* $\widehat{\nu}$*-almost every* $y \in Y^i$ *are* E^i, σ^i*-typical, where* E^i, σ^i *are as in Theorem 4.20.*

Proof. The number of paths in Γ of length n which never enter a component C with $\xi(C) = \lambda$ is $o(\lambda^n)$, so from the definition of $\widehat{\nu}$ it follows that $\widehat{\nu}$-almost every $y \in Y$ enters some C^i. If a path leaves some component with $\xi = \lambda$ it can never enter another one with $\xi = \lambda$, so $\widehat{\nu}$-almost every $y \in Y$ which enter some C^i never leave.

We now prove the second statement of the lemma. The following proof was suggested by Shigenori Matsumoto [**18**].

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 eigenvalue λ. 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 a transition matrix $N(i)$, by restricting μ and N. The measure μ_i is stationary for $N(i)$ (i.e. $\mu_i^{\mathrm{T}} N(i) = \mu_i^{\mathrm{T}}$), so $\widehat{\mu}_i$ is shift invariant. Since C_i is a component, 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. [**20**, 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. 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 §4.2; consequently $q_* \widehat{\nu}_i$ and μ_i are quasi-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

$$\frac{1}{m} \sum_0^m \delta_{S^k \gamma} \to \widehat{\mu}_i.$$

By Theorem 4.20 it follows that the geodesic ray in G associated with any $\gamma \in Y_i^*$ is E^i, σ^i-typical, and the lemma is proved. □

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

Up to this point, we have only used the weak combability of ϕ. The next lemma requires bicombability.

LEMMA 4.23. *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 in ∂G as 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 $d_L(\gamma_i, \gamma_i') \leq C$ and therefore $|\phi(\gamma_i) - \phi(\gamma_i')| \leq K$ for some K independent of i. This shows that γ' is E, σ-typical if γ is. Similarly, if $g \in G$ then $d_R(g\gamma_i, \gamma_i) \leq C$ and therefore $|\phi(g\gamma_i) - \phi(\gamma_i)| \leq K$ for some K independent of i. □

For each component C^i, let $\partial^i G$ denote the image of the E^i, σ^i-typical $y \in Y^i$ under $Y \to \partial G$. Note that $\widehat{\nu}(\partial^i G) > 0$ for all i. By Theorem 4.17 and Lemma 4.16, for any i, j there is $g \in G$ with $\widehat{\nu}(g\partial^i G \cap \partial^j G) > 0$. Hence, by Lemma 4.23 there is an infinite geodesic ray in G which is both E^i, σ^i typical and E^j, σ^j typical. However, this implies $E^i = E^j$ and $\sigma^i = \sigma^j$. Since i and j were arbitrary, we have proved the following result.

LEMMA 4.24. *The means E^i and standard deviations σ^i as above are all equal to some E, σ.*

Recall that the measure $\widehat{\nu}$ on Y pushes forward to a measure on Y_n by projection $Y \to Y_n$ and then to a measure $\widehat{\nu}$ on G_n for all n. Note that $\widehat{\nu}$ on G_n does not depend on the particular choice of a digraph Γ parameterizing L, but just on L itself. Lemma 4.24 and Theorem 4.20 together imply our main result.

THEOREM 4.25. (Central limit theorem for bicombable functions) *Let ϕ be a bicombable function on a word-hyperbolic group with respect to some combing L. 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 $\sigma \geq 0$ where E denotes the mean of $d\phi$ on Γ with respect to the stationary measure μ. In particular, E and σ are algebraic.

Proof. Fix some $g \in G_m$ which corresponds to an element $y \in Y_m$ such that the last vertex $v_g = S^m y \in X_0 = \Gamma$ is in the support of μ. For any $n \geq m$, the shift map S^m takes the subset $p^{-1}(y) \subset Y_n$ with its $\widehat{\nu}$ measure isomorphically to the set of walks in X of length $n - m$ starting at v_g with their μ measure, up to scaling the measures by a constant factor. Fixing g and letting n go to infinity, we obtain the desired Gaussian distribution for the values of ϕ on $\mathrm{cone}(g) \cap G_n$ as $n \to \infty$ in the $\widehat{\nu}$ measure.

For any ϵ, there is m such that the set of $g \in G_m$ with v_g not in the support of μ has $\widehat{\nu}$ measure at most ϵ. Since ϵ was arbitrary, we are done. □

The following corollary does not make reference to the measure $\widehat{\nu}$.

COROLLARY 4.26. *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*

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

the set of elements of length $n \geq N$, there is a subset G'_n with $|G'_n|/|G_n| \geq 1 - \epsilon$, so that for all $g \in G'_n$, there is an inequality

$$|\phi(g) - nE| \leq K \cdot \sqrt{n}.$$

As a special case, let S_1, S_2 be two finite 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, we have the following corollary.

COROLLARY 4.27. *Let S_1 and S_2 be finite generating sets for a word-hyperbolic group G. There is an algebraic number $\lambda_{1,2}$ 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 $n \geq N$, there is a subset G'_n with $|G'_n|/|G_n| \geq 1 - \epsilon$, so that for all $g \in G'_n$ there is an inequality*

$$||g|_{S_1} - \lambda_{1,2}|g|_{S_2}| \leq K \cdot \sqrt{n}.$$

Remark 4.28. A counting argument shows $\lambda_{1,2} \geq \log \lambda_1 / \log \lambda_2$ where the number of words of length n in the S_i metric is $\Theta(\lambda_i^n)$. However, equality should only hold in very special circumstances, and in fact the number $\lambda_{1,2} \log \lambda_2 \log {\lambda_1}^{-1}$ is probably very interesting in general. Such numbers have been studied in free groups in the special case where S_1 and S_2 are free bases, by [**15**].

If ϕ is a quasimorphism, then $|\phi(g) + \phi(g^{-1})| \leq$ constant so if S is symmetric, then necessarily E as above is equal to zero. Hence, we have the following corollary.

COROLLARY 4.29. *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 $\sigma \geq 0$.

5. *Hölder quasimorphisms*

The distribution of quasimorphisms on free groups has been studied by Horsham–Sharp. We briefly describe their results. Let F denote a free group (on some finite generating set).

Definition 5.1. For any $g, a \in F$ and any $\psi : F \to \mathbb{R}$ define

$$\Delta_a \psi(g) = \psi(g) - \psi(ag).$$

For $x, y \in F$ let $(x|y)$ denote the Gromov product

$$(x|y) := (|x| + |y| - |x^{-1}y|)/2.$$

A quasimorphism ψ on 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

$$|\Delta_a \psi(x) - \Delta_a \psi(y)| \leq C e^{-c(x|y)}.$$

Note that the constants C, c depend on a but not on x or y. The main theorem of [**14**], which also appeared in Horsham's PhD thesis, is as follows.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

THEOREM 5.2. (Horsham–Sharp [**14**]) *Let ψ be a Hölder quasimorphism on a free group. If $\overline{\psi}_n$ denotes the value of ψ on a random word in F of length n (with respect to a standard generating set), then there is convergence in the sense of distributions*

$$n^{-1/2}\overline{\psi}_n \to N(0, \sigma)$$

for some σ.

These results can also be generalized to fundamental groups of closed (hyperbolic) surfaces. Big (Brooks) counting quasimorphisms on free groups are Hölder, since $\Delta_a\psi(x) = \Delta_a\psi(y)$ whenever $(x|y)$ is bigger than $|a|$. However, small counting quasimorphisms (i.e. counting quasimorphisms where copies must be disjoint) are not, as the following example shows.

Example 5.3. Let ϕ be the small counting quasimorphism for the word $abab$. Then

$$\phi(\underbrace{bababab\cdots ab}_{4n+1}) = n, \quad \phi(\underbrace{abababab\cdots ab}_{4n+2}) = n$$

but

$$\phi(\underbrace{bababab\cdots ab}_{4n+3}) = n, \quad \phi(\underbrace{abababab\cdots ab}_{4n+4}) = n+1.$$

Of course, since small counting quasimorphisms are bicombable, Theorem 4.25 applies. Note that the measure $\widehat{\nu}$ agrees with the uniform measure in a free group with the standard generating set.

It seems plausible that Horsham–Sharp's methods apply to arbitrary Hölder quasimorphisms on hyperbolic groups, although we have not pursued this.

Acknowledgements. We would like to thank Ian Agol, Ilya Kapovich, Shigenori Matsumoto, Curt McMullen, Mark Pollicott, Richard Sharp, Amie Wilkinson and the anonymous referees for useful comments and corrections. This research was largely carried out while the first author was at the Tokyo Institute of Technology (TIT), hosted by Sadayoshi Kojima. He thanks TIT and Professor Kojima for their hospitality. The second author would like to thank Caltech for their hospitality in hosting him while some of this paper was written. The paper [**19**] was a catalyst for discussions which led to the content of this paper, and we are pleased to acknowledge it. Danny Calegari was supported by NSF grant DMS 0707130. Koji Fujiwara is supported by JSPS-KAKENHI-19340013.

REFERENCES

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

[2] A. Berman and R. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, New York, 1979.

[3] R. Brooks. Some remarks on bounded cohomology. *Riemann Surfaces and Related Topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, NY, 1978) (Annals of Mathematics Studies, 97)*. Princeton University Press, Princeton, NJ, 1981, pp. 53–63.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press

[4] D. Calegari. *scl (MSJ Memoirs, 20)*. Mathematical Society of Japan, Tokyo, 2009.
[5] J. Cannon. The combinatorial structure of cocompact discrete hyperbolic groups. *Geom. Dedicata* **16**(2) (1984), 123–148.
[6] M. Coornaert. Mesures de Patterson–Sullivan sur le bord d'un espace hyperbolique au sens de Gromov. *Pacific J. Math.* **159** (1993), 241–270.
[7] M. Coornaert and A. Papadopoulos. *Symbolic Dynamics and Hyperbolic Groups (Lecture Notes in Mathematics, 1539)*. Springer, Berlin, 1993.
[8] M. Coornaert and A. Papadopoulos. Symbolic coding for the geodesic flow associated to a word hyperbolic group. *Manuscripta Math.* **109** (2002), 465–492.
[9] M. Dehn. Papers on group theory and topology. Translated from the German and with introductions and an appendix by John Stillwell. Springer, New York, 1987.
[10] D. Epstein, J. Cannon, D. Holt, S. Levy, M. Paterson and W. Thurston. *Word Processing in Groups*. Jones and Bartlett Publishers, Boston, MA, 1992.
[11] D. Epstein and K. Fujiwara. The second bounded cohomology of word-hyperbolic groups. *Topology* **36** (1997), 1275–1289.
[12] C. Grinstead and J. Snell. *Introduction to Probability*. American Mathematical Society Publications, Providence, RI, 1997.
[13] M. Gromov. Hyperbolic groups. *Essays in Group Theory (Mathematical Sciences Research Institute Publications, 8)*. Springer, New York, 1987, pp. 75–263.
[14] M. Horsham and R. Sharp. Lengths, quasi-morphisms and statistics for free groups. *Spectral Analysis in Geometry and Number Theory (Dedicated to Toshikazu Sunada on the occasion of his 60th birthday) (Contemporary Mathematics, 484)*. American Mathematical Society, Providence, RI, 2009, pp. 219–237.
[15] V. Kaimanovich, I. Kapovich and P. Schupp. The subadditive ergodic theorem and generic stretching factors for free group automorphisms. *Israel J. Math.* **157** (2007), 1–46.
[16] J. Kemeny and L. Snell. *Finite Markov Chains (University Series in Undergraduate Mathematics)*. Van Nostrand, Princeton, NJ, 1960.
[17] A. Markov. *Extension of Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain*, republished as Appendix B in *Dynamic Probabilistic Systems, Vol. I: Markov Models (Series in Decision and Control, by Richard A. Howard)*. John Wiley and Sons, 1971.
[18] S. Matsumoto. Private communication, 2008.
[19] J.-C. Picaud. Cohomologie bornée des surfaces et courants géodésiques. *Bull. Soc. Math. France* **125** (1997), 115–142.
[20] M. Pollicott and M. Yuri. *Dynamical Systems and Ergodic Theory (London Mathematical Society Student Texts, 40)*. Cambridge University Press, Cambridge, 1998.
[21] S. Rees. Hairdressing in groups: a survey of combings and formal languages. *Geom. Topol. Monogr.* **1** (1998), 493–509.
[22] I. Rivin. Growth in free groups (and other stories). *Preprint*, 1999, math.CO/9911076.
[23] V. Romanovskii. *Discrete Markov chains*. Wolters-Noordhoff Publishing, Groningen, 1970 (translated from the Russian by E. Seneta).
[24] P. Sarnak. Private communication, 2008.
[25] R. Sharp. Local limit theorems for free groups. *Math. Ann.* **321** (2001), 889–904.

https://doi.org/10.1017/S0143385709000662 Published online by Cambridge University Press