CHAPTER 6

# Combable functions and ergodic theory

In this chapter we study quasimorphisms on hyperbolic groups, especially counting quasimorphisms, from a *computational* perspective. We introduce the class of *combable functions* (and the related classes of weakly combable and bicombable functions) on a hyperbolic group, and show that the Epstein–Fujiwara counting functions are bicombable.

Conversely, bicombable function satisfying certain natural conditions are shown to be quasimorphisms; thus quasimorphisms and bounded cohomology arise naturally in the study of automatic structures on hyperbolic groups, a fact which might at first glance seem surprising.

The (asymptotic) distribution of values of a combable function may be described very simply using stationary Markov chains. Consequently, we are able to derive a central limit theorem for the distribution of values of counting quasimorphisms on hyperbolic groups.

The main reference for this section is Calegari–Fujiwara [**50**], although Picaud [**166**] and Horsham–Sharp [**113**] are also relevant.

## 6.1. An example

### 6.1.1. Random walk on $\mathbb{Z}$.

DEFINITION 6.1. A sequence of integers $x = (x_0, x_1, \cdots)$ is a *walk* on $\mathbb{Z}$ if it satisfies the following two properties:

(1) (initialization) $x_0 = 0$
(2) (unit step) for all $n > 0$, there is an equality $|x_n - x_{n-1}| = 1$

The *length* of a walk $x$ is *one less* than the number of terms in the sequence $x$. So, for example, $(0, 1, 2)$ has length 2, while $(0, 1, 0, -1, -2)$ has length 4.

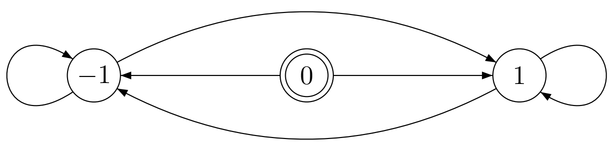


FIGURE 6.1. Walks on $\mathbb{Z}$ of length $n$ are in bijection with walks on $\Gamma$ of length $n$.

Knowing the successive differences $x_n - x_{n-1} \in \{-1, 1\}$ determines $x$, so there is a bijection between walks of length $n$, and strings of length $n$ in the alphabet

$\{-1, 1\}$. This correspondence may be encoded graphically as follows. Let $\Gamma$ be the directed graph depicted in Figure 6.1. A walk $x$ on $\mathbb{Z}$ "determines" a corresponding walk $x'$ on $\Gamma$ starting at the initial vertex (labeled 0) where the labels on the vertices in the itinerary of $x'$ are exactly the sequence of successive differences $x_n - x_{n-1}$. In other words,

$$x'_n = x_n - x_{n-1}$$

Formally, $x'$ is a kind of *discrete derivative* of $x$. The advantage of the correspondence $x \to x'$ is that it replaces a random walk on an infinite (but homogeneous) graph (i.e. $\mathbb{Z}$) with a random walk on a *finite* graph.

Let $X_n$ denote the set of walks on $\mathbb{Z}$ of length $n$, and let $v : X_n \to \mathbb{Z}$ be the function which takes each walk to the last integer in the sequence. For example, $v(0, 1, 2, 1) = 1$ and $v(0, -1, -2, -3, -2) = -2$.

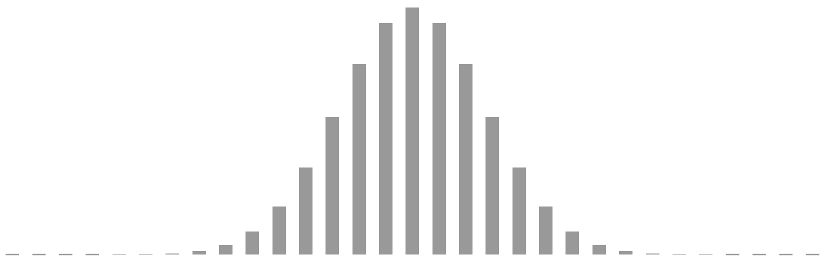

FIGURE 6.2. histogram showing the frequency of outcomes for all walks of length 30 on $\mathbb{Z}$

There are $2^n$ walks of length $n$. The set of values of $v$ on $X_n$ are the integers of the form $2i - n$ for $0 \leq i \leq n$, and the number of elements of $X_n$ taking the value $2i - n$ is $\binom{n}{i} = \frac{n!}{(n-i)!i!}$. A histogram of this data for the case $n = 30$ is contained in Figure 6.2.

This figure has some significant qualitative features: left-right symmetry, the fact that all realized values have the same parity, and so forth. Most notable are the long flat tails on either side. If we rescale the graph horizontally by a factor of $n^{-1}$, and vertically so that the total area under the graph is equal to 1, the distribution becomes more and more peaked and "limits" to a Dirac distribution with all the mass centered at the origin (technically, this is convergence in the sense of distribution). However, if we instead rescale the graph horizontally by a factor of $n^{-1/2}$, the distribution converges to the familiar "bell curve", or Gaussian. If we let $\bar{v}_n$ denote the value of $v$ on a random element of $X_n$ (with the uniform distribution), then $\bar{v}$ is not a number but rather a (discrete) probability measure on $\mathbb{R}$. The Central Limit Theorem for binomial distributions (see [**96**], Thm. 9.1) says that there is convergence in the sense of distribution

$$\lim_{n\to\infty} \mathbf{P}(s \leq n^{-1/2}\bar{v}_n \leq t) = \frac{1}{\sqrt{2\pi}} \int_s^t e^{-x^2/2} dx$$

where $\mathbf{P}(\cdot)$ denotes probability, and $s \leq t$ are any two real numbers.

**6.1.2. Random value of a homomorphism.** Given a group $G$ and a function $f : G \to \mathbb{R}$ it is natural to ask how the values of $f$ are distributed on $G$. If $G$ is finitely generated, we can study *statistical properties* of the values of $f$ on the set of elements of $G$ of (word) length $n$, as a function of $n$. This analysis will be most informative when the function $f$ is adapted to the geometry and algebra of $G$; the most important case therefore is when $f$ is a homomorphism.

In order to keep the discussion concrete, we restrict attention in what follows to free groups. Let $F$ denote the free group generated by two elements $a, b$, and let $\rho : F \to \mathbb{Z}$ be the unique homomorphism which sends $a$ to 1 and $b$ to 0 (writing $\mathbb{Z}$ additively). A basic question is to ask what is the distribution of the values of $\rho$ on the group $F$.

If we take $S = \{a, b, a^{-1}, b^{-1}\}$ to be a symmetric generating set for $F$, the Cayley graph $C_S(F)$ is an infinite regular 4-valent tree. Let $\gamma$ denote a geodesic in $C_S(F)$ starting at id, and let $\rho(\gamma)$ denote the corresponding walk in $\mathbb{Z}$, whose itinerary consists of the values of $\rho$ on successive vertices of $\gamma$. As in the case of a random walk on $\mathbb{Z}$, the situation is clarified by considering, in place of $\rho(\gamma)$, the *discrete derivative*; i.e. by considering how the value of $\rho$ *changes* on successive vertices of $\gamma$.

**6.1.3. Digraphs.** Every element of $F$ is represented by a unique reduced word in the generators, corresponding to the unique geodesic in $C_S(G)$ starting at id and with a given endpoint. Reduced words are certified by *local data*: a word is reduced if and only if no $a$ follows or precedes an $a^{-1}$, and if no $b$ follows or precedes a $b^{-1}$. Let $S^*$ denote the set of all finite words in the generating set $S$, and let $W_n$ denote the set of reduced words in $S^*$ of length $n$. Let $W = \cup_n W_n$.

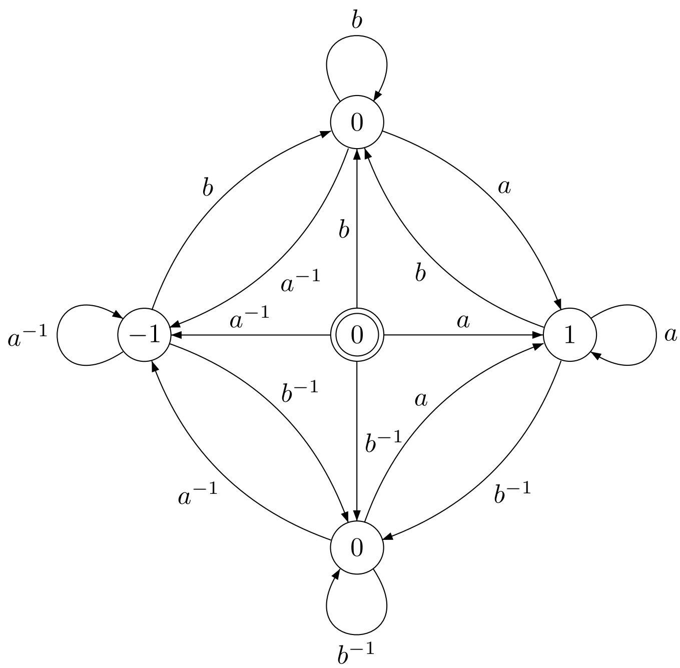


FIGURE 6.3. The digraph $\Gamma$ parameterizes the set of reduced words in $F$

Elements of $F$ are in bijection with elements of $W$ by taking each element to the unique reduced word which represents it. Moreover, elements of $W$ are in bijection with certain walks on a *directed graph* $\Gamma$, depicted in Figure 6.3 (ignore the numbers on the vertices for the moment). There is a special *initial vertex* with no incoming edges, and four other vertices which have both incoming and outgoing edges. In computer science and combinatorics, a directed graph is usually called a *digraph*, and we use this terminology in what follows. If we need to stress that a particular digraph has an initial vertex, we call it a *pointed digraph.* So $\Gamma$ in Figure 6.3 is a (pointed) digraph.

A *reduced* word $w \in W$ determines a directed path in $\Gamma$ starting at the initial vertex, by reading the letters one by one (from left to right) and traversing at each stage the edge of $\Gamma$ labeled by the corresponding letter of $w$. Conversely, a directed path in $\Gamma$ starting at the initial vertex determines a reduced word, determined by the string consisting of the edge labels visited in the path. Under this bijection, elements of $W_n$ correspond to directed paths in $\Gamma$ of length $n$.

The information in a digraph can be encoded in the so-called *adjacency matrix.*

Definition 6.2. Let $\Gamma$ be a digraph with vertices $v_i$. The *adjacency matrix* of $\Gamma$ is the square matrix whose entries are determined by the formula

$$M_{ij} = \begin{cases} 1 \text{ if there is a directed edge from } v_i \text{ to } v_j \\ 0 \text{ otherwise} \end{cases}$$

Spectral properties of $M$ reflect geometric properties of $\Gamma$. The most explicit example of this is the following Lemma, which says that directed paths in $\Gamma$ are counted by the entries of powers of $M$.

Lemma 6.3. *For any $n$ and any vertices $v_i, v_j$ the number of directed paths in $\Gamma$ from $v_i$ to $v_j$ of length $n$ is $(M^n)_{ij}$.*

Proof. We prove the statement by induction. It is tautologically true for paths of length 1, so assume it is true for paths of length $n-1$. By induction, for any $v_k$ there are $(M^{n-1})_{ik}M_{kj}$ paths of length $n$ from $v_i$ to $v_j$ whose penultimate vertex is $v_k$. Summing over $k$ gives the desired result. □

The following topological property of digraphs is the analogue of irreducibility in the algebraic context.

Definition 6.4. A digraph is *recurrent* if there is a directed path from any vertex to any other vertex.

By Lemma 6.3, a digraph is recurrent if and only if for any $i, j$ there is some $n$ (which may depend on $i, j$) for which $(M^n)_{ij}$ is positive.

Definition 6.5. A matrix with non-negative entries is a *Perron–Frobenius matrix* if for any $i, j$ there is some $n$ for which $(M^n)_{ij}$ is positive.

The graph $\Gamma$ of Figure 6.3 is not recurrent, but the subgraph $\Gamma'$ consisting of vertices and edges disjoint from the initial vertex is recurrent. Let $M$ be the adjacency matrix of $\Gamma'$, so

$$M = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

Since $\Gamma'$ is recurrent, the matrix $M$ is a Perron–Frobenius matrix. In fact, every entry of $M^n$ is positive whenever $n \geq 2$.

For such matrices, one has the fundamental Perron–Frobenius Theorem:

THEOREM 6.6 (Perron–Frobenius). *Let $M$ be a real non-negative matrix so that every entry of $M^n$ is positive for some $n > 0$. Then the following statements hold.*

(1) *$M$ has a positive real eigenvalue $\lambda$. Every other eigenvalue $\xi$ satisfies $|\xi| < \lambda$.*
(2) *The algebraic and geometric multiplicities of $\lambda$ are both equal to 1.*
(3) *There are left and right eigenvectors of $M$ with eigenvalue $\lambda$, spanning their respective 1-dimensional eigenspaces, with positive entries.*

See for example [**11**] for a proof. A matrix $M$ with the property above is sometimes called *regular*.

If $M$ is symmetric, the left and right $\lambda$-eigenvectors of $M$ are transposes of each other, but for general $M$ this need not be the case.

For the case of $\Gamma'$ as above, the matrix $M$ is symmetric, and the vector $v = (1/4, 1/4, 1/4, 1/4)$ (resp. $v^T$) is a left (resp. right) eigenvector for $M$ with eigenvalue 3 and $L^1$ norm equal to 1.

If $M$ is merely non-negative (with no assumption that there is a power all of whose entries are strictly positive), the situation is more complicated. Since we will need to study this case in the sequel, we state the following proposition.

PROPOSITION 6.7 (Weak Perron–Frobenius). *Let $M$ be a real non-negative matrix. Then $M$ has a positive real eigenvalue $\lambda$ with left and right eigenvectors, and every other eigenvalue $\xi$ satisfies $|\xi| \le \lambda$.*

*If for every $i, j$ there is an $n$ (possibly depending on $i, j$) for which $(M^n)_{ij}$ is positive, then every eigenvalue $\xi$ with $|\xi| = \lambda$ has the form $\omega\lambda$ for some root of unity $\omega$. Moreover, for every $\xi$ with $|\xi| = \lambda$ the algebraic and geometric multiplicities of $\xi$ are equal, and there are left and right $\lambda$ eigenvectors for $\xi$ with positive entries.*

See [**11**]. A matrix with the property that for all $i, j$ there is $n$ depending on $i, j$ such that $(M^n)_{ij}$ is positive, is sometimes said to be *ergodic* or *irreducible*. An ergodic matrix which is not regular is sometimes called *cyclic*.

To say that the algebraic and geometric multiplicities of an eigenvalue $\xi$ are equal just means that the Jordan block of the eigenvalue $\xi$ is diagonal; i.e. that the generalized $\xi$-eigenspace is a genuine eigenspace. The weak Perron–Frobenius Theorem can be deduced from the (ordinary) Perron–Frobenius Theorem by approximating a non-negative matrix by a positive matrix.

**6.1.4. Random walks on $\Gamma'$.** For each integer $n \ge 0$, let $X_n$ denote the set of walks on $\Gamma'$ of length $n$ starting at any vertex. For each $m < n$ there is a *prefix* function $p_{n,m} : X_n \to X_m$ which just forgets the last $n - m$ terms in the sequence. Each map $X_n \to X_{n-1}$ is finite to one. The inverse limit

$$X := \varprojlim X_n$$

is topologically a Cantor set, and parameterizes the set of right-infinite walks on $\Gamma'$. We write a typical $x \in X_n$ as a finite sequence $x = (x_0, x_1, \cdots, x_n)$ and an element $x \in X$ as an infinite sequence $x = (x_0, x_1, \cdots)$. It comes together with prefix maps $p_n : X \to X_n$ satisfying $p_{n,m} p_n = p_m$ for all $m < n$.

DEFINITION 6.8. The *shift* map $S : X \to X$ takes a walk to the suffix consisting of all but the first vertex. In co-ordinates,

$$S(x_0, x_1, \cdots) = (x_1, x_2, \cdots)$$

DEFINITION 6.9. A *cylinder* is an open subset of $X$ determined by fixing a finite number of the co-ordinates $x_i$ of an element $x$.

Let $\mathcal{B}$ denote the $\sigma$-algebra on $X$ generated by all cylinders. Note that $\mathcal{B}$ is the Borel $\sigma$-algebra on $X$ associated to its natural inverse limit topology.

The shift map $S$ acts continuously on $X$, and therefore measurably with respect to $\mathcal{B}$. Any measurable map on a compact space preserves some probability measure. In our example, there is a unique probability measure $\mu$ on $X$ which is invariant under $S$, with the property that for all $n$, the pushforward $(p_n)_*\mu$ is equal to the uniform probability measure on $X_n$.

If $\pi : X \to \Gamma'$ takes an element to its initial vertex, and $x \in X$ is chosen at random, the sequence

$$\pi(x), \pi(Sx), \pi(S^2x), \cdots$$

is an infinite random walk on $\Gamma'$, where the transition probabilities to move from vertex to vertex at each stage are given by the matrix

$$N = \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{pmatrix}$$

$N$ is a *stochastic matrix*, meaning that the entries are non-negative, and the vector $\mathbf{1} := (1, \cdots, 1)^T$ is a right eigenvector with eigenvalue 1. The uniform probability measure on $\Gamma'$ is stationary for $N$, meaning that $\mathbf{1}^T$ is a left eigenvector with eigenvalue 1. Again, in general, a left eigenvector for a stochastic matrix will *not* correspond to the uniform measure, however a stationary measure exists by the Perron–Frobenius Theorem 6.6.

The essential property of the process $(x_0, x_1, \cdots)$ corresponding to a random $x \in X$ (with respect to the uniform measure) is that for each $i$, the probability that $x_{i+1}$ will be in a given state depends only on $x_i$, and not on $x_j$ for any $j < i$. Informally, this can be summarized by saying that *future states depend only on the present, and are independent of the past.* This property of a random process is generally called the *Markov property*, and the usual terminology for this is that a random walk on $\Gamma'$ is (governed by) a *stationary Markov chain.* The Perron–Frobenius property of the transition matrix $N$ is summarized by saying that this Markov chain is *ergodic.*

For each $n$, let $Y_n$ be the subspace of $X_n$ consisting of walks that begin at the initial vertex. Elements of $Y_n$ are in bijection with elements of $F$ of word length $n$. Each element of $Y_n$ corresponds to a cylinder in $X$ consisting of infinite walks that begin with a given prefix. The measure $\mu$ induces in this way a measure on each $Y_n$; after scaling, this is the uniform measure in which each element has probability $1/(4 \cdot 3^{n-1})$. The homomorphism $\rho$ determines a function $d\rho$ from $\Gamma'$ to $\mathbb{Z}$ by the formula

$$d\rho(\mathbf{s}(ws)) = \rho(ws) - \rho(w)$$

where $\mathbf{s} : Y_n \to \Gamma'$ sends a (finite) walk to its terminal vertex. In other words, the function $d\rho$ measures how much the value of $\rho$ *changes* on the increasing prefixes of a reduced word. If a vertex $v_i$ of $\Gamma'$ is encoded as a column vector, the function $d\rho$ can be encoded as a row vector of the same length, and evaluation of the function amounts to contraction of vectors. In our example, $d\rho$ is the vector $(1, 0, -1, 0)$.

Let $\bar{S}_n$ be a random variable whose value is

$$\bar{S}_n = \sum_{i=1}^{n} d\rho(x_i)$$

where $x = (x_0, x_1, \cdots, x_n)$ is a random element of $Y_n$. In other words, $\bar{S}_n$ is the value of $\rho$ on a random element of $F$ of word length $n$.

Technically, $\bar{S}_n$ should be thought of as a probability measure on $\mathbb{R}$, supported in $\mathbb{Z}$. Every $x \in Y_n$ determines an integer $\sum d\rho(x_i)$, and this determines a map $Y_n \to \mathbb{Z}$. The (uniform) measure on $Y_n$ pushes forward under this map to a measure on $\mathbb{Z}$, which by definition is $\bar{S}_n$.

The Central Limit Theorem for ergodic stationary Markov chains (see [**179**] p. 231) says that there is a convergence in the sense of distribution

$$\lim_{n\to\infty} \mathbf{P}\left(s \le \frac{\bar{S}_n - nE}{\sqrt{\sigma^2 n}} \le t\right) = \frac{1}{\sqrt{2\pi}}\int_s^t e^{-x^2/2}dx$$

where $E$ is the mean of $d\rho$ on $\Gamma'$ with respect to the stationary measure (which is equal to 0 in this case) and $\sigma^2$ is an algebraic number which can be determined from $N$, $\mu$ and $d\rho$.

**6.1.5. More complicated examples.** The homomorphism $\rho$ in the example above is a very simple example of a big counting quasimorphism; explicitly, $\rho = H_a$ in the notation of Definition 2.25. We would like to study the distribution of $H_w$ on $F$ for an arbitrary reduced word $w \in F$. The problem is that the digraph $\Gamma$ defined in the last section is not adequate for our purpose. A reduced word in $F$ determines a walk in $\Gamma$, but the vertex at each step only "remembers" one letter at a time. In order to count occurrences of a word $w$ or its inverse $w^{-1}$ we need a more complicated digraph whose vertices remember enough information to keep track of each occurrence of $w$ or $w^{-1}$.

Definition 6.10. Let $\Gamma$ be a pointed digraph. Define $\Gamma_0 = \Gamma$. For each $n > 0$, define inductively a pointed digraph $\Gamma_n$ as follows.

The vertices of $\Gamma_n$ consist of an initial vertex, together with one vertex for every directed path in $\Gamma_{n-1}$ of length 1 (with any starting vertex). The edges of $\Gamma_n$ (except for those which start at the initial vertex) correspond to pairs of composable paths; i.e. pairs of paths of length 1 which can be concatenated to form a path of length 2.

Finally, for every path of length 1 in $\Gamma_{n-1}$ starting at the initial vertex, add a directed edge in $\Gamma_n$ from the initial vertex to the corresponding vertex of $\Gamma_n$.

$\Gamma_n$ is called the $n$th *refinement* of $\Gamma$.

Remark 6.11. The construction of a refinement makes sense for any pointed digraph.

Remark 6.12. Notice that each $\Gamma_n$ is finite if $\Gamma$ is, and contains a unique maximal recurrent subgraph $\Gamma'_n$ if $\Gamma$ does.

See Figure 6.4 for an example of the first refinement $\Gamma_1$, where $\Gamma$ is the example from Figure 6.3. For the sake of legibility, labels on the arrows (which are elements of the generating set $S$) have been suppressed. Note how complicated this example is, with 17 vertices and 52 edges. In general, the graph $\Gamma_n$ contains $O(\lambda^n)$ vertices and $O(\lambda^{n+1})$ edges, where $\lambda$ is the Perron–Frobenius eigenvalue of the transition matrix of $\Gamma$, so actually constructing $\Gamma_n$ is typically not practical, even for moderate values of $n$.

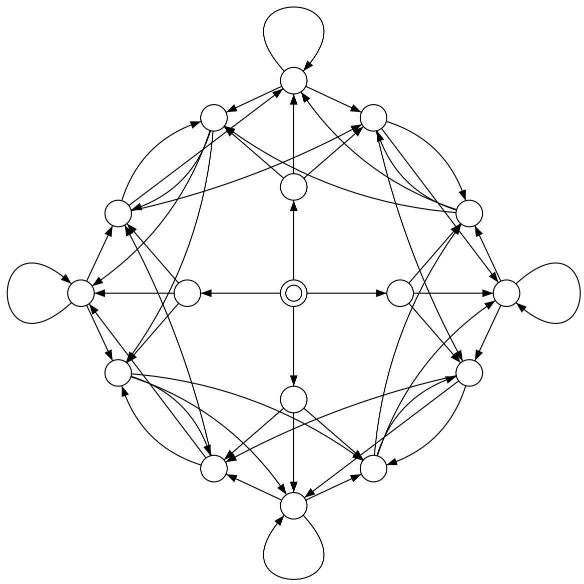

FIGURE 6.4. The digraph $\Gamma_1$ is the (first) refinement of $\Gamma$. $dH_{ab}$ is a function from the states of $\Gamma_1$ to $\mathbb{Z}$.

By induction, the stationary measure for each $\Gamma_n$ is the uniform measure on the subgraph $\Gamma'_n$, and the transition matrix has equal probability for each edge of $\Gamma'_n$.

For any $n, m \geq 0$ there is an equality $(\Gamma_n)_m = \Gamma_{n+m}$. Moreover, by induction, $\Gamma'_n = (\Gamma')_n$. Vertices in $\Gamma'_n$ correspond to paths of length $n$ in $\Gamma'$. Let $w \in F$ be a reduced word of length $n$, and let $H_w = C_w - C_{w^{-1}}$ be the big counting quasimorphism. Define $dH_w : \Gamma'_n \to \mathbb{Z}$ by setting $dH_w$ equal to 1 on the vertex corresponding to the path $w$ in $\Gamma'$, and $-1$ on the vertex corresponding to the path $w^{-1}$ in $\Gamma'$. The Central Limit Theorem for ergodic stationary Markov chains implies the following theorem.

THEOREM 6.13 (Calegari–Fujiwara). *Let $H_w$ be a big counting quasimorphism on a free group. If $\bar{H}_w(n)$ denotes the value of $H_w$ on a random word in $F$ of length $n$ (in a standard symmetric generating set), then there is convergence in the sense of distributions*

$$n^{-1/2}\bar{H}_w(n) \to N(0, \sigma)$$

*for some $\sigma$ depending on $w$.*

It is one of the goals of this chapter to generalize this theorem to a broader class of quasimorphisms on arbitrary word hyperbolic groups.

**6.1.6. Hölder quasimorphisms.** The property of big counting quasimorphisms described in Theorem 6.13 holds for other interesting classes of quasimorphisms on free groups, including those with the so-called *Hölder property.*

DEFINITION 6.14. For any $g \in F$, and any function $\psi$ on $F$, define

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

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

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

In other words, $(x|y)$ is the length of the biggest common prefix of the words $x, y$.

Say that a quasimorphism $\psi \in Q(F)$ is *Hölder* if for any $a \in F$ there are constants $C, c > 0$ such that for any $x, y \in F$ there is an inequality

$$|\Delta_a\psi(x) - \Delta_a\psi(y)| \leq Ce^{-c(x|y)}$$

Note that the constants $C, c$ depend on $a$, but not on $x$ or $y$.

Horsham and Sharp [**113**], extending some results in Matthew Horsham's PhD thesis, prove the following theorem:

THEOREM 6.15 (Horsham–Sharp). *Let $\psi$ be a Hölder quasimorphism on a free group. If $\bar{\psi}(n)$ denotes the value of $\psi$ on a random word in $F$ of length $n$ (in a standard symmetric generating set), then there is convergence in the sense of distributions*

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

*for some $\sigma$.*

The argument involves (nonstationary) Markov chains obtained from subshifts of finite type, and the associated thermodynamic formalism. These results can also be generalized to surface groups.

Big counting quasimorphisms are trivially seen to be Hölder, since $\Delta_a\psi(x) = \Delta_a\psi(y)$ whenever $(x|y)$ is bigger than $|a|$. But small counting quasimorphisms are not, as the following example (from [**50**]) shows.

EXAMPLE 6.16. Let $h := h_{abab}$. Then

$$h(\underbrace{babab\cdots ab}_{4n+1}) = n, \quad h(\underbrace{ababab\cdots ab}_{4n+2}) = n$$

but

$$h(\underbrace{babab\cdots ab}_{4n+3}) = n, \quad h(\underbrace{ababab\cdots ab}_{4n+4}) = n+1$$

Although small counting quasimorphisms are not Hölder, they nevertheless have a great deal in common with big counting quasimorphisms: both are examples of *bicombable functions*, to be defined in § 6.3.2. Ultimately, we will prove a version of the Central Limit Theorem valid for all bicombable functions on arbitrary word-hyperbolic groups.

**6.1.7. Rademacher function.** There are natural ways to filter elements in free groups other than by word length. If one thinks of a (virtually) free group as the fundamental group of a cusped hyperbolic surface (orbifold), it is natural to count conjugacy classes (which correspond to closed geodesics) and sort them by geodesic length. The noncompactness of the surface leads to quite distinctive features of the theory. In this context, we mention a result of Peter Sarnak, showing that the Rademacher function on conjugacy classes in the group $\mathrm{PSL}(2, \mathbb{Z})$ has values which obey a Cauchy distribution, in contrast to the Gaussian distributions discussed above.

Ghys [**91**] gave an elegant topological definition of the Rademacher function. The group $\mathrm{PSL}(2, \mathbb{Z})$ acts on the hyperbolic plane $\mathbb{H}^2$ by isometries, with quotient the $(2, 3, \infty)$-triangle orbifold $\Delta$. Each element $A$ of $\mathrm{PSL}(2, \mathbb{Z})$ whose trace has absolute value $> 2$ fixes a unique axis in $\mathbb{H}^2$, which covers a geodesic in $\Delta$. This geodesic lifts to an embedded loop $\gamma_A$ in the unit tangent bundle $UT\Delta$ which is homeomorphic to the quotient $\mathrm{PSL}(2,\mathbb{R})/\mathrm{PSL}(2,\mathbb{Z})$. As is well known, $UT\Delta$ is homeomorphic to the complement of the trefoil knot $T$ in $S^3$.

DEFINITION 6.17. For $A \in \mathrm{PSL}(2,\mathbb{Z})$ with $|\mathrm{tr}(A)| > 2$, define $R(A)$ to be the linking number of $\gamma_A$ and $T$ in $S^3$.

Ghys relates $R(A)$ to the classical Rademacher function, which is defined in terms of Gauss sums, and is intimately related to the Dedekind $\eta$ function. If we think of $\mathrm{PSL}(2,\mathbb{Z})$ as a subgroup of $\mathrm{Homeo}^+(S^1)$, and $\widetilde{\mathrm{SL}}(2,\mathbb{Z})$ as its preimage in $\mathrm{Homeo}^+(\mathbb{R})^{\mathbb{Z}}$, then there is a rotation quasimorphism rot on $\widetilde{\mathrm{SL}}(2,\mathbb{Z})$. Let $\rho : \widetilde{\mathrm{SL}}(2,\mathbb{Z}) \to \mathbb{Z}$ be the unique homomorphism that takes the value 6 on the generator of the center (i.e. the element that acts on $\mathbb{R}$ as $z \to z+1$). Then $6 \cdot \mathrm{rot} - \rho$ descends to the quasimorphism $R$ on $\mathrm{PSL}(2,\mathbb{Z})$.

Conjugacy classes of elements $A$ in $\mathrm{PSL}(2,\mathbb{Z})$ with $|\mathrm{tr}| > 2$ correspond to closed geodesics $\gamma_A$ in $\Delta$. Let $|\gamma_A|$ denote the length of $\gamma_A$. For each real number $y$, define

$$\pi(y) := \#\{A : |\gamma_A| \le y\}$$

The behavior of $\pi(y)$ for large $y$ is known; in fact,

$$\pi(y) = \mathrm{Li}(e^y) + O(e^{7y/10})$$

where

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x}$$

Sarnak shows that the Rademacher function $R$, filtered by geodesic length, satisfies a Cauchy distribution; i.e.

THEOREM 6.18 (Sarnak). *With notation as above,*

$$\lim_{y\to\infty} \frac{1}{\pi(y)} \#\{A : |\gamma_A| \le y \text{ and } a \le \frac{R(A)}{|\gamma_A|} \le b\} = \frac{1}{\pi}\left(\arctan(\frac{b\pi}{3}) - \arctan(\frac{a\pi}{3})\right)$$

The "reason" for the difference in observed distributions has to do with the relationship between word length and geodesic length in $\mathrm{PSL}(2,\mathbb{Z})$. The group $\mathrm{PSL}(2,\mathbb{Z})$ is virtually free, containing a subgroup $\Gamma$ of index 12 which is isomorphic to $F_2$. The surface $\Delta$ is non-compact, with a cusp. A geodesic $\gamma_A$ which winds a lot around the cusp might have length as small as $O(\log(n))$ where $n$ is the word length of $A$. If $w$ is a reduced word in $F_2$ of the form $a^{n_1}b^{n_2}\cdots a^{n_k}b^{n_k}$ then the length of $w$ is $\sum n_i$ but the length of the geodesic $\gamma_w$ is $O(\sum \log(n_i))$. Since quasimorphisms are homomorphisms on cyclic subgroups, such a word probably has an unusually large value of $R$ for its word length, and especially for its geodesic length, thus giving rise to the fat tails of the Cauchy distribution.

## 6.2. Groups and automata

Our analysis in § 6.1 depended crucially on the fact that elements in a free group could be parameterized by directed paths in a digraph, namely the digraph $\Gamma$ from Figure 6.3 and its refinements. The proper generalization of this fact for more complicated groups involves the theory of combings and regular languages.

**6.2.1. Regular languages.** Let $S$ be a finite alphabet, and let $S^*$ denote the set of all (finite) words in the alphabet $S$.

Definition 6.19. A *language* is a subset $L \subset S^*$. A language is *prefix closed* if every prefix of an element of the language is also in the language.

Definition 6.20. A *finite state automaton* on a fixed alphabet is a digraph with a distinguished initial vertex (the *input state*), and with oriented edges labeled by letters of the alphabet, such that at each vertex there is at most one outgoing edge with any label.

The vertices are also called the *states* of an automaton. A word $w \in S^*$ determines a directed path in the automaton, which starts at the initial vertex at time 0, and moves along a directed edge labeled $w_i$ at time $i$, if one exists, or halts if not. The resulting path in the automaton is said to be obtained by *reading* the word $w$.

Some subset of vertices are labeled *accept* states. If the automaton reads to the end of $w$ without halting, the last vertex of the path is the *final state* and the word is *accepted* if the final state is an accept state, and rejected otherwise.

Definition 6.21. A *regular language* is the set of words in some fixed alphabet accepted by some finite state automaton.

Remark 6.22. For regular languages which are prefix closed, one can restrict attention to automata in which every state is an accept state. In the sequel we shall be exclusively interested in prefix closed regular languages, and therefore every state in our automata will be an accept state.

The concept of a finite state automaton or a regular language is best understood by considering some simple examples.

Example 6.23. Let $S = \{a, b\}$. The following languages are regular:

(1) The set of all words in $S^*$
(2) The set of all words in $S^*$ which contain the string *baa* but not the string *abba*
(3) The set of all words in $S^*$ with at least 5 $a$'s
(4) The set of all words in $S^*$ for which the number of $a$'s and $b$'s have different parities

The following languages are not regular:

(1) The set of all words of the form $a^n b^n$
(2) The set of all palindromic words
(3) The set of all words with prime length
(4) The set of all words which contain more $a$'s than $b$'s

In words, a finite state automaton is a machine with a finite amount of memory. It reads the letters of $w$ in order, and cannot go back and re-read some subword. In practice, automata can be described informally in terms of the task they perform, rather than explicitly in terms of vertices and edges.

Suppose $L$ is regular and prefix closed. Then there is a finite state automaton $\mathsf{A}$ which accepts $L$ and for which every vertex is an accept state. The underlying digraph $\Gamma$ of the automaton $\mathsf{A}$ *parameterizes* $L$, in the sense that there is a natural bijection

$$\text{directed paths in } \Gamma \text{ starting at the initial vertex} \longleftrightarrow \text{ elements of } L$$

NOTATION 6.24. Suppose $\Gamma$ parameterizes $L$. Let $w \in L$ and as above, let $w_i$ denote the $i$th letter of $w$. We let $\gamma_i(w)$ denote the $i$th vertex of the corresponding path in $\Gamma$, respectively $\gamma_i$ if $w$ is understood, and let $\gamma(w)$ (resp. $\gamma$) denote the endpoint of the path in $\Gamma$.

WARNING 6.25. For a fixed regular, prefix closed language $L$ there are *many* digraphs $\Gamma$ which parameterize $L$. For instance, if $\Gamma$ parameterizes $L$, then so does every resolution $\Gamma_n$.

### 6.2.2. Combings.

NOTATION 6.26. If $G$ is a group and $S$ is a generating set, there is a natural evaluation map $\mathsf{e} : S^* \to G$ taking a word in the generators to the element in $G$ it represents. Sometimes, where no confusion can arise, we omit $\mathsf{e}$, so that the same symbol $w$ may represent a word in $S^*$ or an element of $G$. If $w$ is a word in $S^*$, we let $\mathsf{e}_i(w)$ denote the path in $G$ whose $i$th element is the image under $\mathsf{e}$ of the prefix of $w$ of length $i$.

DEFINITION 6.27. Let $G$ be a group with finite symmetric generating set $S$. A *combing* of $G$ with respect to $S$ is a regular language $L \subset S^*$ which satisfies the following conditions:

1. The evaluation map $\mathsf{e} : L \to G$ is a bijection
2. $L$ is prefix closed
3. $L$ is geodesic; i.e. elements of $L$ represent geodesic paths in $C_S(G)$

WARNING 6.28. Definitions of combings differ in the literature. All three bullets in Definition 6.27 (and sometimes even the condition that $L$ is regular) are omitted or modified by some authors!

Let $L$ define a combing of $G$ with respect to $S$, and let $\Gamma$ be a digraph which parameterizes $L$. Every path in $\Gamma$ determines a path in $C_S(G)$ starting at the identity. The conditions in Definition 6.27 imply that the union of these paths is an isometrically embedded maximal spanning tree in $C_S(G)$.

One of the principal motivations for studying combings is the following theorem, first proved by Cannon (though he used different terminology):

THEOREM 6.29 (Cannon [**51**], [**77**]). *Let $G$ be a word-hyperbolic group, and $S$ a finite symmetric generating set. There is a combing of $G$ with respect to $S$.*

In fact, many natural, explicit combings exist. Choose a total ordering $\prec$ on the elements of $S$. This induces a *lexicographic* ordering (i.e. a dictionary ordering) on the elements of $S^*$. The language $L$ of lexicographically first geodesic words in $S^*$ satisfies the bullet conditions of Definition 6.27; the main content of Theorem 6.29 is that $L$ is regular.

## 6.3. Combable functions

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

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

Each metric $d_L, d_R$ is induced from a path metric. The geometry of a metric space $X, d_X$ may be probed effectively by studying the space of all Lipschitz functions $X \to \mathbb{R}$. For $G$ a group, it is natural to probe $G$ by functions which are Lipschitz with respect to either the $d_L$ or $d_R$ metric, or both simultaneously.

Note that a function $f : G \to \mathbb{Z}$ is Lipschitz for the $d_L$ metric if and only if there is a constant $C$ so that for all $a \in G$ and all $s \in S$,

$$|f(as) - f(a)| \le C$$

Similarly, $f$ is Lipschitz for the $d_R$ metric if

$$|f(sa) - f(a)| \le C$$

The properties of being Lipschitz for $d_L$ or $d_R$ respectively do not depend on a choice of generating set for $S$ (but the constants will).

Remark 6.30. It is psychologically challenging to find a good way to perceive a group $G$ simultaneously in both its $d_L$ and $d_R$ metrics. An analogy is the relationship between matrices and rooted trees. The elements of a matrix can be thought of as the leaves of a depth 2 rooted tree in two distinct ways. The depth 1 nodes can either be thought of as denoting *rows* or as *columns*. The two tree structures are obtained by thinking of the index sets as affine spaces for the action of a group $\mathbb{Z}$, and the two different tree structures correspond to the actions of $\mathbb{Z}$ from the left and from the right.

Any homomorphism $G \to \mathbb{Z}$ is Lipschitz in both the $d_L$ and $d_R$ metrics. But hyperbolic groups do not always admit many (or even any) homomorphisms to $\mathbb{Z}$ (for instance, fundamental groups of quaternionic hyperbolic manifolds have Kazhdan's property (T), and therefore no subgroup of finite index admits a homomorphism to $\mathbb{Z}$). However, quasimorphisms are also obviously Lipschitz in both the $d_L$ and $d_R$ metrics, and therefore any hyperbolic group is guaranteed a rich family of such functions.

**6.3.2. Combable functions.** We now introduce the class of *combable functions* on a hyperbolic group $G$.

Definition 6.31. Let $G$ be word-hyperbolic with finite symmetric generating set $S$, and let $L \subset S^*$ be a combing of $G$ with respect to $S$. A function $\phi : G \to \mathbb{Z}$ is *weakly combable* with respect to $S$, $L$ (or *weakly combable* if $S$, $L$ are understood) if there is a digraph $\Gamma$ parameterizing $L$ and a function $d\phi$ from the vertices of $\Gamma$ to $\mathbb{Z}$, such that for any word $w \in L$ there is an equality

$$\phi(\mathsf{e}(w)) = \sum_i d\phi(\gamma_i(w))$$

(here $\mathsf{e}(w)$ on the left denotes an element of $G$ and $\gamma_i(w)$ on the right denotes the vertices in $\Gamma$ of the path corresponding to $w \in L$). If the maps $\mathsf{e}$ and $\gamma$ are understood, by abuse of notation we write this formula as

$$\phi(w) = \sum_i d\phi(w_i)$$

A function $\phi$ is *combable* if it is weakly combable and is Lipschitz as a map from $G, d_L \to \mathbb{Z}$. It is *bicombable* if it is weakly combable and is Lipschitz both as a map from $G, d_L \to \mathbb{Z}$ and from $G, d_R \to \mathbb{Z}$.

A weakly combable function is *ergodic* (resp. almost ergodic) if there is an automaton $\Gamma$ parameterizing $L$ which has a unique maximal recurrent subgraph (resp. with maximal eigenvalue), and is *regular* if it is ergodic, and its recurrent subgraph is aperiodic.

WARNING 6.32. Remember that a combing $L$ with respect to $S$ can be parameterized by many different graphs $\Gamma$. If $\phi$ is weakly combable with respect to $S, L$ then there is *some* digraph $\Gamma$ parameterizing $L$ for which $d\phi$ is a function on $\Gamma$. The particular parameterizing digraph $\Gamma$ may definitely depend on $\phi$.

REMARK 6.33. There is no strict logical necessity to restrict attention to functions with values in $\mathbb{Z}$. One can vary the definition and for any finitely generated group $H$ define weakly combable $H$-functions, by defining $d\phi : \Gamma \to H$ and replacing sum by group multiplication in $H$. Since $H$ is finitely generated, it makes sense to talk about left Lipschitz and right Lipschitz functions from $G$ to $H$ and therefore to define combable and bicombable $H$-functions. Notice with this definition that any homomorphism $G \to H$ is a bicombable $H$-function.

EXAMPLE 6.34. Word length is bicombable.

REMARK 6.35. Theorem 6.29 remains true, and with essentially the same proof, when $S$ is an *asymmetric* generating set which generates $G$ as a semigroup. For semigroup generators, one must slightly change the definition of a combing to say that words in $L$ represent shortest *directed* paths to their endpoints, rather than geodesics in $C_S(G)$. It follows that Example 6.34 remains true in the more general context of word length with respect to an asymmetric set of generators for $G$ (as a semigroup).

The definition of weakly combable depends quite strongly on the choice of the generating set $S$, as the following example shows.

EXAMPLE 6.36. Let $G = \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$, and let the factors be generated by $a$ and $b$ respectively. Define $f : G \to \mathbb{Z}$ by

$$f(w) = \begin{cases} n \text{ if } w = a^n \text{ for some } n \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

Then $f$ is weakly combable with respect to the generating set $a, a^{-1}, b$; a digraph to calculate $f$ is depicted in Figure 6.5. On the other hand, $f$ is not weakly combable

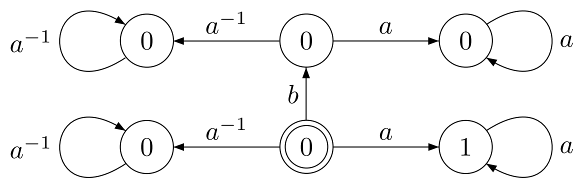


FIGURE 6.5. A digraph to calculate $f$

with respect to the generating set $ab, a^{-1}b, b$. Note for this generating set that $(ab)^n$ is the unique geodesic representing its value in $G$, and therefore $(ab)^n \in L$ for all $n$. Suppose to the contrary that $f$ is weakly combable with respect to $ab, a^{-1}b, b$, so there is a finite digraph $\Gamma$ parameterizing $L$ and a function $df : \Gamma \to \mathbb{Z}$ as in Definition 6.31. Since $\Gamma$ is finite, there is a constant $C$ such that $|f(w) - f(ws)| \leq C$ whenever $w$ and $ws$ are words in $L$ which differ by right multiplication by a single generator. Yet $f((ab)^{2n}) = 2n$ for $n \geq 0$, whereas $f((ab)^{2n+1}) = 0$, so no such pair $L, \Gamma$ can exist.

Example 6.36 shows that the property of weak combability is contingent, and perhaps not so useful. By contrast, the Independence Theorem (Theorem 6.39, to be proved shortly) shows that combability is *independent* of the choice of generating set. For this reason, combable functions are much more useful and interesting than weakly combable functions.

We introduce some definitions which will be useful in what follows.

DEFINITION 6.37. Let $G$ be hyperbolic with finite symmetric generating set $S$. Let $L$ be a combing with respect to $S$. Let $B$ be the ball of radius $N$ about id in $G$ with the metric inherited from $C_S(G)$, and let $\Sigma$ be a finite set. A *tile set* is a map

$$T : B \times G \to \Sigma$$

such that for any pair of words $w, ws \in L$ where $s \in S$, the map $T(\cdot, \mathsf{e}(ws)) : B \to \Sigma$ depends only on $T(\cdot, \mathsf{e}(w))$ and $s$.

DEFINITION 6.38. Let $T$ be a tile set, and let $\Gamma$ be a digraph parameterizing $L$. The *fiber product* is the digraph $\Gamma_T$ parameterizing $L$ defined as follows. The vertices of $\Gamma_T$ are the functions of the form

$$(T(\cdot, \mathsf{e}(w)), \gamma(w)) : B \to \Sigma \times \Gamma$$

and $(T(\cdot, \mathsf{e}(w)), \gamma(w))$ is joined to $(T(\cdot, \mathsf{e}(ws)), \gamma(ws))$ by an edge labeled $s$ whenever $w, ws \in L$.

Geometrically, $\Gamma_T$ can be thought of as a bundle over $\Gamma$ whose fiber at each vertex $v$ is the (finite) set of functions of the form $T(\cdot, \mathsf{e}(w)) : B \to \Sigma$ for all $w$ satisfying $\gamma(w) = v$.

THEOREM 6.39 (Independence of combability). *Let $\phi : G \to \mathbb{Z}$ be combable with respect to some $S'$, $L'$. Then for any other generating set $S$ and any combing $L$ with respect to $S$, the function $\phi$ is combable with respect to $S$, $L$.*

PROOF. If $S, S'$ are two generating sets, and $L, L'$ are two bijective geodesic combings, then every word in $L'$ is quasigeodesic in $C_S(G)$ and by the Morse Lemma (Theorem 3.30, bullet (1)), (asynchronously) fellow travels the word in $L$ with the same evaluation. That is, there are constants $N$ and $k$ such that the following is true:

(1) For all words $w'$ in $L'$ and $w$ in $L$ with $\mathsf{e}(w') = \mathsf{e}(w)$, the path $w'$ (i.e. the set of $\mathsf{e}_i(w')$) is contained in the $N$ neighborhood of the path $w$ (i.e. the set of $\mathsf{e}_i(w)$) in $C_S(G)$. Furthermore, the path $w'$ intersects the $N$ neighborhood of *every* vertex on $w$ (i.e. it comes uniformly close to every vertex on $w$)
(2) If $\mathsf{e}_i(w') \in B_N(\mathsf{e}_j(w))$ and $\mathsf{e}_l(w') \in B_N(\mathsf{e}_{j+1}(w))$, then $|l - i| < k$

Bullet (1) may be restated informally as saying that for every $w' \in L'$ and $w \in L$ with $\mathsf{e}(w') = \mathsf{e}(w)$, the path corresponding to $w'$ is obtained by concatenating paths $x_i$ of uniformly bounded length, whose endpoints are within a bounded distance of successive vertices of $w$.

Now, suppose $\phi$ is combable with respect to $L'$. Let $\Gamma'$ be a digraph which parameterizes words in $L'$ for which $d\phi : \Gamma' \to \mathbb{Z}$ is defined. Let $B$ denote the ball of radius $N$ around id in $G$ with the metric inherited from $C_S(G)$.

We define a tile set $T$ taking values in a certain finite set as follows. For each $g \in G$ and $h \in B$, let $w \in L$ and $w' \in L'$ evaluate to $g$ and $gh$. That is, $\mathsf{e}(w) = g$

and $\mathsf{e}(w') = gh$. If some $\mathsf{e}_i(w')$ is not contained in the $N$ neighborhood of any $\mathsf{e}_j(w)$, or if the $N$ neighborhood of some $\mathsf{e}_j(w)$ does not intersect $w'$ (i.e. if the conditions of bullet (1) above are violated), then $T(h,g) = \mathtt{E}$, an "out of range" symbol. Otherwise set

$$T(h,g) = (\phi(gh) - \phi(g), \gamma(w'))$$

in other words, the tuple consisting of the difference of $\phi$ on $gh$ and $g$, and the vertex of $\Gamma'$ corresponding to the endpoint of the path $w'$.

In words, for a fixed $g \in G$, the set of pairs $h, g$ parameterizes the ball of radius $N$ about $g$. For every element $gh$ of this ball, there is a unique path in $L'$ which evaluates to $gh$. If this path does not stay in the $N$ neighborhood of the path in $L$ evaluating to $g$, the value of $T$ is out of range. Otherwise, $T$ calculates the value of $\phi$ on the element $gh$ (normalized by subtracting the value of $\phi$ on $g$) and the vertex of $\Gamma'$ associated to the word of $L'$ corresponding to $gh$.

Since $\phi$ is Lipschitz in the $C_{S'}(G)$ metric, it is also Lipschitz in the $C_S(G)$ metric, so the normalized values of $\phi$ on $B_N(g)$ are uniformly bounded, independent of $g \in G$. This shows that $T$ takes values in a *finite* set. This is the only place where combability (as distinct to weak combability) is used in the proof. We will show that $T$ is a *tile set*.

REMARK 6.40. In fact, the second factor of $T$ is by itself already a tile set; on a first reading, it is worth verifying this fact alone, and then seeing how it can be used to deduce the stronger claim about $T$.

To verify that $T$ is a tile set, we just need to check that if $w, ws \in L$ then $T(\cdot, \mathsf{e}(ws))$ depends only on $T(\cdot, \mathsf{e}(w))$ and on $s$.

Let $h \in B$, and suppose $w' \in L'$ is such that $\mathsf{e}(w') = \mathsf{e}(ws)h \in G$. If the path $w'$ is contained in the $N$ neighborhood of the path $ws$, there is a factorization $w' = v'x$ in $L'$ where $\mathsf{e}(v')$ is within distance $N$ of $\mathsf{e}(w)$, and where $x$ is a path in $\Gamma'$ of length $\le k$. So for each $f \in B$ with $\mathsf{e}(v') = \mathsf{e}(w)f$ we can enumerate the set of all paths $\alpha$ in $\Gamma'$ of length $\le k$ starting at $\gamma(v')$, and see whether $f\mathsf{e}(\alpha) = \mathsf{e}(s)h$. If no such $f, \alpha$ exists, then $T(h, \mathsf{e}(ws)) = \mathtt{E}$. Otherwise, the state $\gamma(w')$ can be deduced from the state $\gamma(v')$ and from $x$ (this shows that the second factor of $T$ is a tile set), and we can calculate

$$\phi(\mathsf{e}(ws)h) - \phi(\mathsf{e}(w)f) = \sum_i d\phi(\alpha_i)$$

If $h = \mathrm{id}$ then some such $f, \alpha$ is guaranteed to exist, by the discussion above. Hence $\phi(\mathsf{e}(ws)) - \phi(\mathsf{e}(w)f)$ can be calculated, and therefore for any $h \in B$ we can calculate $\phi(\mathsf{e}(ws)h) - \phi(\mathsf{e}(ws))$ without using $w$, and therefore $T(\cdot, \mathsf{e}(ws))$ depends only on $T(\cdot, \mathsf{e}(w))$ and on $s$, not on $ws$. This shows that $T$ is a tile set.

If $\Gamma$ is a digraph parameterizing $L$, we build the fiber product $\Gamma_T$. Since $\phi(\mathsf{e}(ws)) - \phi(\mathsf{e}(w)f)$ and $\phi(\mathsf{e}(w)f) - \phi(\mathsf{e}(w))$ depend only on $T(\cdot, w)$ and $s$, the value of $\phi(\mathsf{e}(ws)) - \phi(\mathsf{e}(w))$ depends only on $s$ and the vertex $\gamma(w)$ of $\Gamma_T$. So we can define $d\phi$ as a function on the resolution $(\Gamma_T)_1$ of $\Gamma_T$, where the value of $d\phi$ on the vertex of $(\Gamma_T)_1$ corresponding to the edge from $\gamma(w)$ to $\gamma(ws)$ is equal to $\phi(\mathsf{e}(ws)) - \phi(\mathsf{e}(w))$.

By construction, $d\phi$ satisfies

$$\phi(\mathsf{e}(w)) = \sum_i d\phi(\gamma_i(w))$$

and therefore $\phi$ is combable with respect to $S, L$. $\square$

NOTATION 6.41. Denote the class of combable and bicombable functions on $G$ by $\mathfrak{C}(G)$ and $\mathfrak{B}(G)$ respectively.

LEMMA 6.42. *$\mathfrak{C}(G)$ and $\mathfrak{B}(G)$ are free Abelian groups.*

PROOF. If $\phi$ is (bi-)combable, then obviously so is $-\phi$.

Let $\phi_1, \phi_2$ be combable. Then they are combable with respect to some fixed combing $S, L$. Let $\Gamma_1, \Gamma_2$ be digraphs parameterizing $L$ for which $d\phi_i : \Gamma_i \to \mathbb{Z}$ is defined. Define a new digraph $\Gamma$ with one vertex for each pair of vertices from $\Gamma_1, \Gamma_2$ and with an edge labeled $s$ from $(v_1, v_2)$ to $(v_1', v_2')$ if and only if there is an edge of $\Gamma_i$ from $v_i$ to $v_i'$ labeled $s$ for $i = 1, 2$. The initial vertex of $\Gamma$ is the pair consisting of the initial vertices of $\Gamma_1, \Gamma_2$ respectively. Let $\Gamma'$ be the subgraph of $\Gamma$ consisting of the union of all directed paths starting at the initial vertex. Then $\Gamma'$ parameterizes $L$, and $d(\phi_1 + \phi_2)$ is a function on $\Gamma'$ defined by

$$d(\phi_1 + \phi_2)(v_1, v_2) = d\phi_1(v_1) + d\phi_2(v_2)$$

and therefore $\phi_1 + \phi_2$ is weakly combable. A sum of two functions which are Lipschitz in the $d_L$ (resp. $d_R$) metric is Lipschitz in the $d_L$ (resp. $d_R$) metric, so $\phi_1 + \phi_2$ is (bi-)combable if both $\phi_i$ are.

This shows that $\mathfrak{C}(G)$ and $\mathfrak{B}(G)$ are Abelian groups. Since they take values in $\mathbb{Z}$, they are torsion-free, and not infinitely divisible. $\square$

EXAMPLE 6.43. Let $G = F_2 = \langle a, b \rangle$ and let $f : G \to \mathbb{Z}$ be defined by

$$f(w) = \begin{cases} |w| \text{ if } w \text{ starts with } a \\ 0 \text{ otherwise} \end{cases}$$

Then $f$ is weakly combable; a digraph to calculate $f$ is illustrated in Figure 6.6.

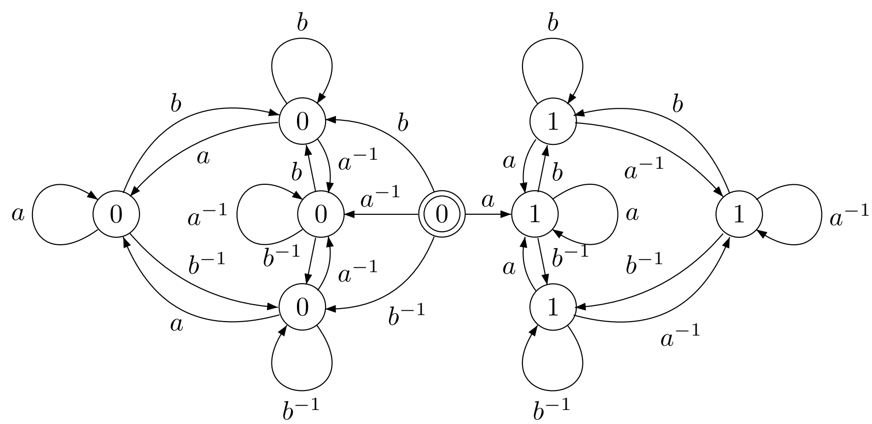


FIGURE 6.6. A digraph to calculate $f$

Moreover, $f$ is Lipschitz in the $d_L$ metric and therefore combable. However, $f(a^n) = n$ whereas $f(ba^n) = 0$ so $f$ is not Lipschitz in the $d_R$ metric, and is not bicombable.

**6.3.3. Quasimorphisms.** There are several natural operations which can be defined on functions $\phi : G \to \mathbb{R}$, including the following:

(1) The *adjoint* of $\phi$, denoted $\phi^*$, defined by

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

(2) The *antisymmetrization* of $\phi$, denoted $\phi'$, defined by

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

In general, neither operation preserves weak combability, although if both $\phi$ and $\phi^*$ are weakly combable, so is $2\phi'$.

LEMMA 6.44. *Suppose $\phi$ is weakly combable, and Lipschitz in the $d_R$ metric. Then there is a constant $C$ so that if $w \in L$ is expressed as a product of subwords $w = uv$ then $|\phi(w) - \phi(u) - \phi(v)| \le C$.*

PROOF. Let $\Gamma$ be a digraph which parameterizes $L$. Let $u' \in L$ be any word such that $\gamma(u) = \gamma(u')$. Then $\phi(w) = \phi(u) + \phi(u'v) - \phi(u')$. Choose $u'$ so that $|u'| \le |\Gamma|$. Since $\phi$ is Lipschitz in the $d_R$ metric, there is a constant $C_1$ so that $|\phi(u'v) - \phi(v)| \le C_1$. Since $|u'|$ is bounded, there is a constant $C_2$ so that $|\phi(u')| \le C_2$. Hence

$$|\phi(w) - \phi(u) - \phi(v)| \le |\phi(u'v) - \phi(u') - \phi(v)| \le C_1 + C_2$$

proving the Lemma. □

In words, Lemma 6.44 says that $\phi$ is almost additive under decomposition.

LEMMA 6.45. *Suppose $\phi$ is bicombable. Then there is a constant $C$ so that if $w \in L$ is expressed as a product of subwords $w = uv$ then*

$$|\phi^*(w) - \phi^*(u) - \phi^*(v)| \le C$$

PROOF. We have $w^{-1} = v^{-1}u^{-1}$ in $G$ but not necessarily in $L$. Let $z \in L$ represent $w^{-1}$, and express $z$ as a product of subwords $z = xy$ where $d_L(v^{-1}, x) \le \delta$ and $d_R(u^{-1}, y) \le \delta$. By Lemma 6.44, $|\phi(z) - \phi(x) - \phi(y)| \le C$. But $\phi(z) = \phi^*(w)$ whereas $|\phi^*(v) - \phi(x)| \le \delta C_1$ and $|\phi^*(u) - \phi(y)| \le \delta C_1$ for some $C_1$ because $\phi$ is bicombable (and therefore Lipschitz in both $d_L$ and $d_R$). The Lemma now follows from the triangle inequality. □

THEOREM 6.46. *Let $\phi : G \to \mathbb{Z}$ be bicombable. Then the antisymmetrization $\phi'$ is a quasimorphism.*

PROOF. Let $u, v \in L$ be arbitrary, and let $w \in L$ satisfy $\mathsf{e}(w) = \mathsf{e}(u)\mathsf{e}(v)$. Then we can write $u = u'x$, $v = yv'$ and $w = w_1w_2$ as words in $L$ so that $d_L(y, x^{-1}) \le \delta$, $d_L(u, w_1) \le \delta$ and $d_R(v, w_2) \le \delta$, by $\delta$-thinness of triangles in $C_S(G)$.

Now apply Lemma 6.44, Lemma 6.45 and antisymmetry. □

REMARK 6.47. Say that a function $\phi : G \to \mathbb{R}$ is *almost antisymmetric* if there is a constant $C$ so that $|\phi(a) + \phi^*(a)| \le C$ for all $a \in G$. The arguments above can be modified to show that an almost antisymmetric bicombable function is a quasimorphism.

Theorem 6.46 can be used to give a surprisingly simple construction of non-trivial quasimorphisms on any hyperbolic group.

EXAMPLE 6.48. Let $G$ be hyperbolic, and let $T$ be a finite *asymmetric* set which generates $G$ as a semigroup. Let $w_T : G \to \mathbb{Z}$ be word length with respect to $T$. Then define

$$h_T(a) = w_T(a) - w_T(a^{-1})$$

for all $a \in G$.

By Remark 6.35, $w_T$ is bicombable and therefore $h_T$ is a quasimorphism.

In fact, it is straightforward to give a direct proof that $h_T$ is a quasimorphism,. Let $S$ be the symmetrization of $T$, and construct the Cayley graph $C_S(G)$. First of all, it is obvious that $h_T$ is Lipschitz in both the $d_R$ and the $d_L$ metrics.

Secondly, every word in $T$ is a path in $C_S(G)$ (but not conversely). A shortest path in $C_S(G)$ from id to $a$ representing a word in $T$ will be called a *realizing path* for $a$. Since every element in $S$ can be written as a word of bounded length in $T$, there are uniform constants $k, \epsilon$ so that realizing paths are $k, \epsilon$ quasigeodesic in $C_S(G)$. In particular, if $l_a$ and $l_{a^{-1}}$ are realizing paths for $a$ and $a^{-1}$ respectively, then $l_a$ and $al_{a^{-1}}$ are $\delta'$ close for some $\delta'$ not depending on $a$. So if $u$ is arbitrary, and $u = vw$ where $v$ is on a realizing path for $u$, then $w^{-1}$ is within distance $\delta'$ of a realizing path for $u^{-1}$. It follows that there is a constant $C$ such that

$$|h_T(u) - h_T(v) - h_T(w)| \leq C$$

for any such factorization. In other words, $h_T$ is almost additive under decomposition.

Now, if $a, b$ are arbitrary, and $l_a, l_b, l_{ab}$ are realizing paths for $a, b, ab$ respectively, then $l_a, al_b, l_{ab}$ are three sides of a $\delta'$ thin quasigeodesic triangle. This triangle can be decomposed into six segments which are $\delta'$ close in pairs. Since $h_T$ is antisymmetric, and Lipschitz in both $d_L$ and $d_R$, the values of $h_T$ on paired segments almost cancel. Since $h_T$ is almost additive under decomposition, $h_T(ab)$ and $h_T(a) + h_T(b)$ are almost equal, and we are done. This shows that $h_T$ is a quasimorphism.

For typical asymmetric $T$, the function $h_T$ is unbounded. This is not completely trivial, but follows from estimates on the length of anti-aligned translates of an axis (compare with Remark 3.12). When $G$ is nonelementary, by varying the choice of generating sets $T$ and taking infinite ($L_1$) linear combinations, one can construct a subspace of $Q(G)$ with dimension $2^{\aleph_0}$, giving a new proof of the main theorem of Epstein–Fujiwara ([**78**], Thm. 1.1).

## 6.4. Counting quasimorphisms

**6.4.1. Greedy algorithm.** In § 3.5 we discussed Fujiwara's construction of counting quasimorphisms associated to an action of a group $G$ on a $\delta$-hyperbolic graph $X$. In the special case that $G$ is a hyperbolic group, and $X$ is the Cayley graph of $G$ with respect to a finite generating set $S$, such quasimorphisms were constructed first by Epstein–Fujiwara [**78**], generalizing Brooks [**27**]. Our aim in this section and the next is to show that counting quasimorphisms are bicombable.

For the sake of clarity, we spell out the definition of Epstein–Fujiwara counting quasimorphisms.

DEFINITION 6.49. Let $G$ be a hyperbolic group with symmetric generating set $S$. Let $\sigma$ be an oriented simplicial path in the Cayley graph $C_S(G)$ and let $\sigma^{-1}$ denote the same path with the opposite orientation. For $\gamma$ an oriented simplicial path in $C_S(G)$, let $|\gamma|_\sigma$ denote the maximal number of disjoint copies of $\sigma$ contained in $\gamma$. For $a \in G$, define

$$c_\sigma(a) = \text{dist}(\text{id}, a) - \inf_\gamma(\text{length}(\gamma) - |\gamma|_\sigma)$$

where the infimum is taken over *all* directed paths $\gamma$ in $C_S(G)$ from id to $a$.

Define a (small) *counting quasimorphism* to be a function of the form

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

This is a special case of Fujiwara's construction in § 3.5 and therefore when $|\sigma| \geq 2$, Lemma 3.46 applies, and realizing paths are (uniformly) quasigeodesic in $C_S(G)$.

Let $\sigma$ be a string. If $w$ is a word, let $|w|_\sigma$ count the maximal number of disjoint copies of $\sigma$ in $w$. Similarly, let $|w|'_\sigma$ count disjoint copies of $\sigma$ in $w$ using the *greedy algorithm*. In other words, define $|w|'_\sigma$ inductively on the length of $w$ by the equality

$$|w|'_\sigma = |v|'_\sigma + 1$$

where $v$ is the word obtained from $w$ by deleting the prefix up to and including the first occurrence of $\sigma$ in $w$.

The advantage of $|\cdot|'_\sigma$ over $|\cdot|_\sigma$ is that it is evident from the definition that $d|\cdot|'$ can be calculated by a finite state automaton. On the other hand, we have the following:

LEMMA 6.50 (Greedy is good). *The functions $|\cdot|_\sigma$ and $|\cdot|'_\sigma$ are equal.*

PROOF. Suppose not, and let $w$ be a shortest word such that $|w|_\sigma$ and $|w|'_\sigma$ are not equal. By definition, $|w|'_\sigma < |w|_\sigma$, and since $w$ is the shortest word with this property, by comparing the values of the two functions on prefixes of $w$, we conclude $|w|'_\sigma = |w|_\sigma - 1$. Since $w$ is the shortest word with this property, the suffix of $w$ must be a copy of $\sigma$ that is counted by $|\cdot|_\sigma$ but not by $|\cdot|'_\sigma$. Hence the greedy algorithm must count a copy of $\sigma$ that overlaps this suffix. Deleting the terminal copy of $\sigma$ reduces the values of both $|\cdot|_\sigma$ and $|\cdot|'_\sigma$ by 1, contrary to the hypothesis that $w$ was shortest. □

### 6.4.2. Counting quasimorphisms are bicombable.

THEOREM 6.51 (Calegari–Fujiwara [**50**]). *Let $G$ be hyperbolic, and let $h_\sigma$ be an Epstein–Fujiwara counting quasimorphism. Then $h_\sigma$ is bicombable.*

PROOF. We give a somewhat informal proof, which can be made rigorous by translating it into the language of tile sets, and following the model of Theorem 6.39.

Fix a hyperbolic group $G$ and a symmetric generating set $S$. Let $L$ be a combing for $G$. Remember that this means that $L$ is a prefix-closed regular language of geodesics in $G$ (with respect to the fixed generating set $S$) for which the evaluation map is a bijection $L \to G$. If $w \in L$ corresponds to $\overline{w}$ in $G$, let $\gamma_w$ be the path in $C_S(G)$ from id to $\overline{w}$.

Let $\sigma$ be a string. We will show that both $c_\sigma$ and $c_{\sigma^{-1}}$ are weakly combable with respect to the generating set $S$ and any combing $L$. By bullet (2) of Lemma 3.45, these functions are Lipschitz in the $d_L$ metric, and therefore combable. Lemma 6.42 implies that their difference $h_\sigma$ is also combable; since it is a quasimorphism, it is bicombable.

In the remainder of the proof, for the sake of clarity, we abbreviate $c_\sigma$ to $c$.

Fix a word $w \in L$. By Lemma 3.46, a realizing path $\alpha$ for $\overline{w}$ is a $K, \epsilon$ quasigeodesic, and therefore by the Morse Lemma, there is a constant $N$ depending only on $\delta, K, \epsilon$ (and not on $w$) so that $\alpha$ and $\gamma_w$ are contained in $N$-neighborhoods of each other. Hence every vertex of $\alpha$ is contained in the $N$-neighborhood of some vertex of $\gamma_w$ and conversely. For each $i$, let $B_N(\gamma_w(i))$ denote the $N$-neighborhood of $\gamma_w(i)$. By uniform quasigeodesity of $\alpha$, and geodesity of $\gamma$, if $p \in B_N(\gamma_w(i))$ and $q \in B_N(\gamma_w(i+1))$ are both on $\alpha$, then the segment of $\alpha$ from $p$ to $q$ has uniformly bounded length. Let $p \in B_N(\gamma_w(i))$ for some $i$. Say a path $\gamma'$ from id to $p$ is *admissible* if it is $K, \epsilon$-quasigeodesic, and if for all $j < i$ the path $\gamma'$ intersects $B_N(\gamma_w(j))$. Thus, an admissible path is obtained by concatenating paths of bounded length whose endpoints are contained in $N$-neighborhoods of successive vertices of $\gamma_w$.

For each $p \in B_N(\gamma_w(i))$ and each path $\gamma'$ from id to $p$, recall that $|\gamma'|$ is the maximal number of disjoint copies of $\sigma$ in $\gamma'$. By Lemma 6.50, the greedy algorithm picks out $|\gamma'|$ specific disjoint copies which we refer to as the *greedy copies* of $\sigma$ in $\gamma'$; let $\sigma(\gamma')$ be the biggest prefix of $\sigma$ which is a suffix of $\gamma'$ and which is *disjoint* from the greedy copies of $\sigma$ in $\gamma'$. Let $X$ denote the set of possible values of $\sigma(\gamma')$. Note that $|X| = |\sigma|$, since the values of $X$ are in bijection with proper prefixes of $\sigma$. One can think of the set $X$ as the states of an automaton that reads a word, and finds the greedy copies of $\sigma$ in that word.

We define a function $T$ as follows. The domain of $T$ is $B_N(\text{id}) \times X \times G$. Fix $h \in B_N(\text{id})$ and $\gamma_w(i) \in G$. Let $g = \gamma_w(i)h \in B_N(\gamma_w(i))$. For each $x \in X$, consider the set of all admissible paths $\gamma'$ from id to $g$ that satisfy $\sigma(\gamma') = x$. If no such path exists, define $T(h, x, \gamma_w(i)) = \mathtt{E}$, an "out of range" symbol. Otherwise, define

$$c(g, x) = \text{dist}(\text{id}, g) - \inf_{\gamma'}(\text{length}(\gamma') - |\gamma'|)$$

where the infimum is taken over $\gamma'$ as above. Notice that $\max_x c(g, x) = c(g)$ if there is some admissible realizing path. In particular, $\max_x c(\gamma_w(i), x) = c(\gamma_w(i))$. If some $\gamma'$ exists as above, define

$$T(h, x, \gamma_w(i)) = c(\gamma_w(i)) - c(g, x)$$

If there is any admissible path $\gamma'$ from id to $g$ that ends in state $x$, there is such a path obtained by composing a realizing path for $\gamma_w(i)$ with a suffix of bounded length. Together with bullet (2) of Lemma 3.45, this implies that $T$ takes values in a finite set.

Suppose we know the value of $T$ on $B_N(\text{id}) \times X \times \gamma_w(i)$. Let $h \in B_N(\text{id})$, and define $g' = \gamma_w(i+1)h$. Any admissible path from id to $g'$ is obtained by concatenating an admissible path from id to some $g \in B_N(\gamma_w(i))$ with a path of bounded length. So if we can compute $d(\text{id}, g') - d(\text{id}, g)$ we can compute $c(g', x) - c(g, y)$ for any $x, y \in X$. Since $G$ is hyperbolic, and $\gamma_w$ is geodesic, we can keep track of relative distances from id to points in the ball of radius $N$ about points on $\gamma_w$, and therefore we can compute $d(\text{id}, g') - d(\text{id}, g)$ by keeping track of only a finite amount of information at each stage. We define a digraph parameterizing $L$ that keeps track at each stage of the following two pieces of information, thought of as functions on the ball of radius $N$ about the current vertex in $C_S(G)$:

(1) The relative distances from id
(2) The value of $T$

By the discussion above, this is a finite digraph, and $dc$ is well-defined as a function on the vertices of its first refinement (cf. Theorem 6.39). Hence $c_\sigma$ and $c_{\sigma^{-1}}$ are combable, and the proof follows. $\square$

REMARK 6.52. The pair consisting of $T$ and relative distance to id is almost a tile set, except that the domain is slightly larger (since $T$ depends, in addition to $B_N(\mathrm{id})$ and $G$, on the choice of an element in the finite set $X$). Otherwise, the proof is conceptually very similar to that of Theorem 6.39.

## 6.5. Patterson–Sullivan measures

The crucial difficulty in extending Theorem 6.13 to general word-hyperbolic groups is the fact that the digraphs associated to arbitrary word-hyperbolic groups are (typically) not recurrent. This means that the stationary Markov chains obtained by generalizing the construction of § 6.1.4 are not typically ergodic, and the Perron–Frobenius Theorem (i.e. Theorem 6.6) does not directly apply.

The first important result we use in this section is Coornaert's Theorem, which says that in a non-elementary word-hyperbolic group $G$, if we fix a finite generating set, there are constants $\lambda > 1$ and $K \geq 1$ so that the number of words of length $n$ is bounded between $K^{-1}\lambda^n$ and $K\lambda^n$ for all $n$. This implies that one can find a digraph parameterizing a combing of $G$ which is *almost semisimple* — that is, the eigenspace of largest absolute value is *diagonalizable*, and the system (measurably) decomposes into a finite number of independent ergodic subsystems. Consequently, most long geodesics in $G$ can be partitioned into finitely many families, each (more-or-less) parameterized by random walks on a *recurrent* digraph whose associated stationary Markov chain is ergodic, and obeys a central limit theorem.

*A priori*, there is no apparent way to compare long geodesics in different families. However, in place of recurrence of a single digraph, one can use the ergodicity of the action of $G$ *at infinity*, on the boundary $\partial G$ with its Patterson–Sullivan measure. A typical infinite geodesic in one family can be translated by left-multiplication to within bounded distance of a typical infinite geodesic in any other family. A bicombable function is almost invariant under both left and right multiplication by elements of bounded size, so the distribution of values on a typical infinite geodesic in one family is *the same* as the distribution on a typical infinite geodesic in the other. In other words, the values of the function on typical paths in one family have the same distribution as the values of typical paths in any other, and we obtain a central limit theorem for the group as a whole. The next few sections flesh out the details of this scheme.

**6.5.1. Some linear algebra.** Let $\Gamma$ be a finite pointed digraph. Let $V$ be the real vector space spanned by the vertices of $\Gamma$, and let $\langle\cdot,\cdot\rangle$ be the inner product on $V$ for which the vertices are an orthonormal basis.

The vertices of $\Gamma$ are denoted $v_i$ for $i \in \{1, \cdots, n\}$. We let $v_1$ denote the initial vertex. For a vector $v \in V$, let $|v|$ denote the $L^1$ norm of $v$. That is,

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

For brevity, let $\mathbf{1}$ denote the vector with all co-ordinates equal to 1, so for a non-negative vector $v$, there is equality $|v| = \langle v, \mathbf{1}\rangle$.

The digraphs $\Gamma$ that parameterize combings of hyperbolic groups are not completely general, but satisfy a number of special properties. We formalize these properties as follows. Let $M$ denote the adjacency matrix of $\Gamma$, so that the number of directed paths in $\Gamma$ of length $n$ from $v_i$ to $v_j$ is

$$(v_i)^T M^n v_j = \langle v_i, M^n v_j \rangle = (M^n)_{ij}$$

DEFINITION 6.53. A digraph $\Gamma$ 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 $\Gamma$ from $v_1$ to $v_i$
(3) There are constants $\lambda > 1, K \geq 1$ so that
$$K^{-1}\lambda^n \leq |v_1^T M^n| \leq K\lambda^n$$
for all positive integers $n$

In what follows we will assume that $\Gamma$ is almost semisimple.

LEMMA 6.54. *Suppose $\Gamma$ is almost semisimple. Then $\lambda$ is the largest real eigenvalue of $M$. Moreover, for every eigenvalue $\xi$ of $M$ either $|\xi| < \lambda$ or else the geometric and the algebraic multiplicities of $\xi$ are equal.*

PROOF. It is convenient to work with $M^T$ in place of $M$. To prove the lemma, it suffices to prove analogous facts about the matrix $M^T$. Corresponding to the Jordan decomposition of $M^T$ over $\mathbb{C}$, let $\xi_1, \ldots, \xi_m$ be the eigenvalues of the corresponding Jordan blocks (listed with multiplicity).

Bullet (2) from Definition 6.53 implies that for any $v_i$, there is an inequality $|(M^n)^T v_i| \leq C_i|(M^n)^T v_1|$ for some constant $C_i$. Since the $v_i$ span $V$, and since $V$ is finite dimensional, there is a constant $C$ such that for all $w \in V$ there is an inequality $|(M^n)^T w| \leq C|(M^n)^T v_1||w|$.

For each $i$, there is some $w_i$ in the $\xi_i$-eigenspace for which

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

where $k$ is the dimension of the Jordan block associated to $\xi_i$. Since $|(M^n)^T w_i| \leq C|(M^n)^T v_1||w_i|$, by bullet (3) from Definition 6.53, either $|\xi| < \lambda$ or $|\xi| = \lambda$ and $k = 1$.

By the Perron–Frobenius theorem for non-negative matrices, $M^T$ has a largest real eigenvalue $\lambda'$ such that $|\xi| \leq \lambda'$ for all eigenvalues $\xi$. We must have $\lambda' = \lambda$ by the estimates above. Note that $M$ has the same spectrum as $M^T$ with the same multiplicity, and that all the $\xi$ eigenspaces of $M$ are diagonalizable for $|\xi| = \lambda$. □

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

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

and similarly for $M^T$.

LEMMA 6.55. *For any vector $v \in V$, the following limit*

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

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

PROOF. We suppress $v$ in the notation that follows. For each eigenvector $\xi$ define

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

And set $\rho_n = \sum_\xi \rho_n(\xi)$. With this notation, $\rho = \lim_{n\to\infty} \rho_n$, and we want to show that this limit exists.

By Lemma 6.54, for each $\xi$, either $|\xi| < \lambda$ or $v(\xi)$ is a $\xi$-eigenvector. In the first case, $\rho_n(\xi) \to 0$. In the second case, either $\xi = \lambda$, or else the vectors $\lambda^{-i}M^i v(\xi)$ become equidistributed in the unit circle in the complex line of $V \otimes \mathbb{C}$ spanned by $v(\xi)$. It follows that $\rho_n(\xi) \to 0$ unless $\xi = \lambda$.

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

Since every eigenvalue of $M$ with largest (absolute) value has geometric multiplicity equal to its algebraic multiplicity, the same is true of the transpose $M^T$. The same argument as Lemma 6.55 implies

LEMMA 6.56. *For any vector $v \in V$, the following limit*

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

*exists, and $(\ell(v))^T$ is the projection of $v^T$ onto the left $\lambda$ eigenspace of $M$.*

For any $v_i$, the partial sums $\rho_n(v_i)$ are non-negative real vectors so if $v$ is non-negative, so is $\rho(v)$. Similarly, if $v$ is non-negative, so is $\ell(v)$.

PROPOSITION 6.57. *For any $v, w \in V$ there is equality*

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

PROOF. By definition,

$$\begin{aligned}
\langle \ell(v), \rho(w)\rangle &= \lim_{n\to\infty} \left( n^{-1} \sum_{i\le n} \lambda^{-i} v^T M^i \right) \left( n^{-1} \sum_{j \le n} \lambda^{-j} M^j w \right) \\
&= \lim_{n\to\infty} n^{-2} \sum_{i,j\le n} \lambda^{-i-j} v^T M^{i+j} w \\
&= \lim_{n\to\infty} n^{-2} \sum_{k \le 2n} (n+1-|n-k|)\lambda^{-k} v^T M^k w \\
&= \lim_{n\to\infty} n^{-2} \sum_{k \le 2n} (n+1-|n-k|)\lambda^{-k} \ell(v)^T M^k w \\
&= \lim_{n\to\infty} n^{-2} n(n+1) \ell(v)^T w = \langle \ell(v), w\rangle
\end{aligned}$$

where the third last equality follows from the "almost periodicity" of $\lambda^{-1}M$ so that all terms except the (left and right) $\lambda$-eigenvalues cancel over any long consecutive sequence of indices. We get $\langle \ell(v), \rho(w)\rangle = \langle v, \rho(w)\rangle$ by the same reason. □

Recall that a *component* of $\Gamma$ is a maximal recurrent subgraph $C$; i.e. a subgraph with the property that there is a directed path from any vertex to any other vertex. Each component $C$ has its own adjacency matrix, with biggest real eigenvalue $\xi(C)$. Since $C$ is a subgraph of $\Gamma$, we must have $\xi(C) \le \xi(\Gamma) = \lambda$ for any $C$.

LEMMA 6.58. *Let $\Gamma$ be almost semi-simple. If $C, C'$ are distinct components with $\xi(C) = \xi(C') = \lambda$ then there is no directed path from $C$ to $C'$.*

PROOF. Recall the Landau notation $f(x) = \Theta(g(x))$ if the ratio $f(x)/g(x)$ is bounded away from zero and away from infinity.

Let $u$ be a vertex in $C$ and $v$ a vertex in $C'$ such that there is a directed path $\gamma$ from $u$ to $v$. Since $C$ is recurrent, Proposition 6.7 implies that there are $\Theta(\lambda^n)$ directed paths in $C$ starting at $u$ of length $n$, and similarly for paths in $C'$ starting at $v$. There is a constant $k$ so that each vertex in $C$ can be joined by a path of length at most $k$ to some $v$. So for each pair of integers $i, n-i$ consider the set of paths of length between $n$ and $n+k$ which consist of an initial segment of length $i$ in $C$ starting at $u$, followed by a path of length $\leq k$ to $v$, followed by a terminal segment of length $n-i$ in $C$. The number of such paths for fixed $i$ is $\Theta(\lambda^n)$, so the number of paths for varying $i$ is $\Theta(n\lambda^n)$. But if $\Gamma$ is almost semi-simple, the number of paths of length between $n$ and $n+k$ (of any kind) is $\Theta(\lambda^n)$, so we obtain a contradiction. $\square$

**6.5.2. Coornaert's Theorem and Patterson–Sullivan measures.** Let $G$ be a non-elementary word-hyperbolic group with generating set $S$. For $g \in G$, let $|g|$ denote word length with respect to $S$.

DEFINITION 6.59. The *Poincaré series* of $G$ is the series

$$\zeta_G(s) = \sum_{g \in G} e^{-s|g|}$$

This series diverges for all sufficiently small $s$, and converges for all sufficiently large $s$. The *critical exponent* is the supremum of the values of $s$ for which the series diverges. Similar zeta functions appear in many contexts, for example in number theory and dynamics. The best results can be expected when the series diverges at the critical exponent.

THEOREM 6.60 (Coornaert, [**56**] Thm. 7.2). *Let $G$ be a non-elementary word-hyperbolic group with generating set $S$. Let $G_n$ be the set of elements of word length $n$. Then there are constants $\lambda > 1, K \geq 1$ so that*

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

*for all positive integers $n$.*

It follows from Theorem 6.60 that the critical exponent of the Poincaré series is equal to $\log(\lambda)$, and the series $\zeta_G(\log(\lambda))$ diverges.

For each $n$, let $\nu_n$ 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 $\delta_g$ is the Dirac measure on the element $g$. The measure $\nu_n$ extends trivially to a probability measure on the compact space $G \cup \partial G$, where $\partial G$ denotes the ideal (Gromov) boundary of $G$.

DEFINITION 6.61. A weak limit $\nu := \lim_{n\to\infty} \nu_n$ is a *Patterson–Sullivan* measure associated to $S$.

Since the Poincaré series diverges at the critical exponent, the support of $\nu$ is contained in $\partial G$.

It is convenient, for the sake of computations, to work with a slightly different normalization of $\nu$. For each $n$, let $\widehat{\nu}_n$ be the measure on $G$ defined by

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

and let $\widehat{\nu} := \lim_{n\to\infty} \widehat{\nu}_n$ be a weak limit. Of course the measures $\widehat{\nu}_n$ and $\nu_n$ are proportional for each $n$. Moreover, by Theorem 6.60, the constant of proportionality is bounded above and below.

REMARK 6.62. In fact, the limit of the $\widehat{\nu}_n$ exists and is well-defined. This is guaranteed by an explicit formula for $\widehat{\nu}$, which is given in § 6.5.3.

The group $G$ acts on itself by left-multiplication. This action extends continuously to a left action $G \times \partial G \to \partial G$. Patterson–Sullivan measures enjoy a number of useful properties, summarized in the following theorem.

THEOREM 6.63 (Coornaert, [**56**] Thm. 7.7). *Let $\nu$ be a Patterson–Sullivan measure. The action of $G$ on $\partial G$ preserves the measure class of $\nu$. Moreover, the action of $G$ on $(\partial G, \nu)$ is ergodic.*

The meaning of ergodicity for a group action which preserves a measure class but not a measure is that for any $A, B \subset \partial G$ with positive $\nu$-measure, there is $g \in G$ with $\nu(gA \cap B) > 0$. Since $\nu$ and $\widehat{\nu}$ are proportional, the action of $G$ on $\partial G$ is also ergodic for the $\widehat{\nu}$ measure.

In fact, Coornaert proves the stronger fact that there is a constant $K > 1$ so that for any $s \in S$ there is an inequality

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

and the same is true for the measure $\widehat{\nu}$, though we do not use this stronger fact.

**6.5.3. Construction of stationary measure.** Throughout the sequel we fix the following notation.

Let $G$ be word-hyperbolic, and $\phi : G \to \mathbb{Z}$ a bicombable function. Fix a finite generating set $S$, and let $L \subset S^*$ be a combing of $G$ with respect to $S$. Since $\phi$ is bicombable, $d\phi$ exists as a map from $\Gamma \to \mathbb{Z}$ for some digraph $\Gamma$ parameterizing $L$, by Theorem 6.39.

Let $M$ denote the adjacency matrix of $\Gamma$, acting on $V$, the space of real-valued functions on the vertices of $\Gamma$. Let $v_1 \in V$ be the function taking the value 1 on the initial vertex, and 0 on all other vertices. Let $\mathbf{1}$ denote the constant function taking the value 1 on every vertex of $\Gamma$.

For each $n$ let $X_n$ denote the set of walks of length $n$ on $\Gamma$ (starting at an arbitrary vertex) and $Y_n$ the set of walks of length $n$ starting at the initial vertex. There are restriction maps $X_{n+1} \to X_n$ and $Y_{n+1} \to Y_n$ for each $n$, with inverse limits $X$ and $Y$. Evaluation of words gives rise to bijections $Y_n \to G_n$ for all $n$; taking limits, there is a map $Y \to \partial G$, called the *endpoint map*, taking an infinite word to the endpoint of the corresponding geodesic ray in $G$.

LEMMA 6.64 (Coornaert–Papadopoulos). *The endpoint map $Y \to \partial G$ is surjective, and bounded-to-one.*

See [**57**] for a proof.

REMARK 6.65. In fact, obtaining a bound on the size of the preimage of a point in $\partial G$ is straightforward. If $\gamma, \gamma'$ are infinite geodesics corresponding to paths in $Y$ with the same endpoint in $\partial G$, then their Hausdorff distance is bounded by $\delta$, the constant of hyperbolicity of $G$. For any point $\gamma_i \in \gamma$, let $B_i$ denote the ball of radius $\delta$ about $\gamma_i$. Then $\gamma'$ must intersect $B_i$, and the prefix of $\gamma'$ up to this point of intersection is uniquely determined by the fact that $\gamma'$ corresponds to a path in $Y$. Hence $\gamma'$ may be thought of as an element of the inverse limit of a partially defined system of maps $B_i \to B_{i-1}$. Since $|B_i| \le C$ for all $i$ for some constant $C$, the cardinality of this inverse limit is also bounded by $C$.

Each $g \in G_n$ corresponds to a unique word $w \in L$ and a unique path $y \in Y_n$. For each $m > n$ the projection $p : Y_m \to Y_n$ determines a subset $p^{-1}(y) \in Y_m$ and a corresponding subset of $G_m$. The set of $h \in G$ corresponding to words $z$ in some $Y_m$ which restricts to a fixed $y$ is called the *cone* of $g$, and denoted $\mathrm{cone}(g)$. Note that $\mathrm{cone}(g)$ depends on $L$, but not on $\Gamma$. For each fixed $n$, we can define a measure $\widehat{\nu}$ on $G_n$ by

$$\widehat{\nu}(g) = \lim_{m\to\infty} \widehat{\nu}_m(\mathrm{cone}(g))$$

(an explicit formula is given below). Identifying $G_n$ with $Y_n$, we obtain a measure on $Y_n$ for each $n$ which by abuse of notation we denote $\widehat{\nu}$. Observe that these measures for different $n$ have the following compatibility property: for each $y \in Y_n$ and each $m > n$, there is an equality $\widehat{\nu}(p^{-1}(y)) = \widehat{\nu}(y)$ where $p : Y_m \to Y_n$ is the restriction map. This compatibility property means that we can define a measure $\widehat{\nu}$ on $Y$ by the formula

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

where $p : Y \to Y_n$ is restriction. Since the cylinders $p^{-1}(y)$ generate the Borel $\sigma$-algebra of $Y$, this defines a unique measure $\widehat{\nu}$ on $Y$ which by construction pushes forward under $Y \to \partial G$ to the measure $\widehat{\nu}$ of the same name on $\partial G$.

We can obtain an explicit formula for the value of $\widehat{\nu}$ on an element $g \in G_n$ or the corresponding element $y \in Y_n$ or cylinder $p^{-1}(y) \subset Y$. By definition, for any $g \in G_n$ and any $m \ge n$ we have

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

Let $v_g \in \Gamma$ be the last vertex of $y$. Then we can rewrite this formula as

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

and therefore by taking limits $m \to \infty$ we obtain the formula

$$\widehat{\nu}(\overline{\mathrm{cone}(g)}) = \lambda^{-n}|\ell(v_g)| = \lambda^{-n}\langle \ell(v_g), \mathbf{1}\rangle = \lambda^{-n}\langle v_g, \rho(\mathbf{1})\rangle$$

where overline denotes closure in $G \cup \partial G$, and where we have used the property that $\ell(\cdot)$ of a non-negative vector is non-negative, and Proposition 6.57 for the last equality.

The measure $\widehat{\nu}$ on $Y$ is typically not invariant under the shift map $S : X \to X$. In fact, $S(Y) \cap Y = \emptyset$ if the initial vertex has no incoming edges. We define a measure $\mu$ on $X$ by

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

and observe that the result is manifestly invariant by $S$. Using the explicit formula for $\widehat{\nu}$ on $Y_n$ and $Y$ we can derive an explicit formula for $\mu$, showing that $\mu$ is well-defined.

Let $v_j \in \Gamma = X_0$ be an arbitrary vertex. By abuse of notation, we let $p : X \to X_0$ denote the restriction map of an infinite path to its initial vertex (this is similar to, but should not be confused with, the restriction maps $p : Y_m \to Y_n$ discussed earlier). We will calculate $\mu(p^{-1}(v_j))$. For each $n$ we can calculate

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

On the other hand, the number of $y \in Y_n$ with $S^n y = v_j$ is exactly equal to the number of directed paths in $\Gamma$ of length $n$ which end at $v_j$, which is $\langle v_1, M^n v_j\rangle$. It follows that

$$\begin{aligned}\mu(p^{-1}(v_j)) &= \lim_{n\to\infty} \langle v_j, \rho(\mathbf{1})\rangle\langle v_1, \frac{1}{n}\sum_{i\le n} \lambda^{-n} M^n v_j\rangle \\ &= \langle v_j, \rho(\mathbf{1})\rangle\langle v_1, \rho(v_j)\rangle = \langle v_j, \rho(\mathbf{1})\rangle\langle \ell(v_1), v_j\rangle\end{aligned}$$

If we define a measure $\mu$ on $\Gamma$ by $\mu_i = \rho(\mathbf{1})_i \ell(v_1)_i$ (where subscripts denote vector components) then it follows that the map $X \to \Gamma$ taking each walk to its initial vertex pushes forward the measure $\mu$ on $X$ to the measure $\mu$ on $\Gamma$.

Define a matrix $N$ with entries

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

if $\rho(\mathbf{1})_i$ is nonzero, and set $N_{ii} = 1$ and $N_{ij} = 0$ otherwise. Recall that a non-negative matrix $N$ with the property that $\sum_j N_{ij} = 1$ for any $i$ is called a *stochastic matrix* (compare with the matrix $N$ in § 6.1.4).

LEMMA 6.66. *The matrix $N$ is stochastic, and satisfies $\mu N = \mu$.*

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

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

This shows $N$ is a stochastic matrix. To verify the second formula,

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

where the sum is over $i$ with $\mu_i \ne 0$ which implies $\rho(\mathbf{1})_i \ne 0$. $\square$

By a further abuse of notation, we let $p : X \to X_n$ denote the restriction of an infinite path to a suitable prefix. We can obtain a formula for the measure $\mu$ on cylinders $p^{-1}(x) \subset X$ for $x \in X_n$ in terms of the measure $\mu$ on $\Gamma$, and the matrix $N$.

LEMMA 6.67. *For $x \in X_n$, there is equality*

$$\mu(p^{-1}(x)) = \mu_{i_0} N_{i_0 i_1} N_{i_1 i_2} \cdots N_{i_{n-1} i_n}$$

*where $x = (x_{i_0}, x_{i_1}, \cdots, x_{i_n})$, and $x_{i_j}$ corresponds to the vertex $v_{i_j}$ of $\Gamma$.*

PROOF. Let $g \in G_n$. If $\gamma_g$ is the corresponding walk in $\Gamma$, let $v_i$ be the last vertex of $\gamma_g$. Then $\widehat{\nu}(g) = \lambda^{-n}\rho(\mathbf{1})_i$. Moreover, for each vertex $v_j$, there are $M_{ij}$ elements $h \in \text{cone}(g)$ for which the corresponding walks $\gamma_h$ have last vertex $v_j$. Each $h$ has $\widehat{\nu}(h) = \lambda^{-n-1}\rho(\mathbf{1})_j$ so given $g$, the sum over $h \in G_{n+1}$ with $h \in \text{cone}(g)$ which have last vertex $v_j$ of $\widehat{\nu}(h)$ is $M_{ij}\rho(\mathbf{1})_j/\lambda\rho(\mathbf{1})_i = N_{ij}$. In other words, given any $y \in Y$ whose $n$th vertex is $v_i$, the probability in the $\widehat{\nu}$ measure that its $(n+1)$st vertex is $v_j$ is $N_{ij}$. Since this formula does not depend on $n$ but just $v_i$ and $v_j$, the lemma is proved. □

We call $\mu$ on $\Gamma$ the *stationary measure*. It is not necessarily a probability measure, but it determines a unique probability measure by scaling. By abuse of notation, we refer to these two measures by the same name. Lemma 6.67 may be interpreted as saying that a random walk on $\Gamma$ with initial vertex chosen randomly with respect to the stationary measure $\mu$ and with transition probabilities given by the stochastic matrix $N$ agrees with a random element of $X$ with respect to the measure $\mu$.

The next Lemma describes the support of the stationary measure $\mu$ on $\Gamma$.

LEMMA 6.68. *The support of the stationary measure is equal to the disjoint union of the maximal recurrent subgraphs $C^i$ of $\Gamma$ whose adjacency matrices have biggest eigenvalue $\lambda$.*

PROOF. Since $\mu_i = \rho(\mathbf{1})_i \ell(v_1)_i$ a vertex $v_i$ is in the support of $\mu_i$ if for some large fixed $k$ there are $\Theta(\lambda^n)$ paths of length between $n$ and $n+k$ from $v_1$ to $v_i$, and $\Theta(\lambda^n)$ paths of length $n$ from $v_i$ to some other vertex. It follows that some path from $v_1$ to $v_i$ intersects a maximal recurrent component $C$ whose adjacency matrix has biggest real eigenvalue $\xi(C) = \lambda$, and similarly there is some outgoing path from $v_i$ which intersects a maximal recurrent component $C'$ with $\xi(C') = \lambda$. Lemma 6.58 implies that $C = C'$, and therefore $v_i \in C$.

Conversely, let $C$ be a recurrent subgraph of $\Gamma$ whose adjacency matrix has eigenvalue $\lambda$. Then $\rho(\mathbf{1})_i$ and $\ell(v_1)_i$ are positive for all $v_i$ in $C$, by counting only paths which stay in $C$ outside a prefix and suffix of bounded length. □

From the point of view of stationary measure, $\Gamma$ decomposes into a finite union of recurrent subgraphs $C^i$, each with Perron–Frobenius eigenvalue $\lambda$. Let $N|_{C^i}$ denote the restriction of the stochastic matrix $N$ to the subgraph $C^i$. Then $N|_{C^i}$ is a stochastic matrix. Let $\mu^i$ denote the measure $\mu$ on $\Gamma$ restricted to $C^i$, and rescaled to be a probability measure. Then $N|_{C^i}$ preserves $\mu^i$, and determines an *ergodic stationary Markov chain* on the vertices of $C^i$.

Let $\phi$ be weakly combable. As in § 6.1.4 we can define $\overline{S}_n^i$ to be equal to the sum of the values of $d\phi$ on a random walk on $C^i$ of length $n$ with respect to the stationary measure $\mu^i$ and transition probabilities given by $N|_{C^i}$.

The Central Limit Theorem for ergodic stationary Markov chains implies

LEMMA 6.69. *Let $\phi$ be weakly combable. With terminology as above, there is convergence in the sense of distribution*

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

*for some $\sigma^i \geq 0$ where $E^i$ denotes the average of $d\phi$ on $C^i$ with respect to the stationary measure $\mu^i$, and $N(0,\sigma^i)$ denotes the Gaussian normal distribution with mean* 0 *and standard deviation $\sigma^i$.*

This theorem is essentially due to Markov [**146**]. For a proof and more details, as well as a precise formula for $\sigma$, see e.g. [**179**], Chapter 4, § 46 or [**96**], § 11.5, especially Theorem 11.17. An excellent general reference is [**127**].

REMARK 6.70. Note that $\sigma = 0$ is possible (for instance, $\phi$ could be identically zero), in which case by convention, $N(0,\sigma)$ denotes the Dirac distribution with mass 1 centered at 0.

**6.5.4. Central Limit Theorem.** In order to derive a central limit theorem for the group $G$ as a whole, we must compare the means $E^i$ and standard deviations $\sigma^i$ associated to distinct components $C^i$.

For each component $C^i$ in the support of the stationary measure $\mu$, let $Y^i \subset Y$ denote the set of infinite paths in $\Gamma$ which eventually enter $C^i$ and stay there. Note that the $Y^i$ are disjoint, and $\widehat{\nu}(Y - \cup_i Y^i) = 0$. For each path $\gamma \in Y$ we can consider the following. Let $\gamma_i \in G$ be the element corresponding to the evaluation of the word which is equal to the prefix of $\gamma$ of length $i$. We fix the following notation: if $r$ is a real number, let $\delta(r)$ denote the probability measure on $\mathbb{R}$ which consists of an atom concentrated at $r$. For a given real number $A$, and for integers $n, m$ we can consider the following measure.

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

and then define $\omega(\gamma) = \lim_{n\to\infty}\lim_{m\to\infty}\omega(n,m)(\gamma)$. Note that the existence of this limit depends on the "correct" choice of $A$.

DEFINITION 6.71. Let $\gamma \in Y^i$. We say that $\gamma$ is *typical* if $\omega(\gamma)$ exists for $A = E^i$, and is equal to $N(0,\sigma^i)$. More generally, if $\gamma$ is an infinite geodesic ray in $G$, then $\gamma$ is $E,\sigma$-typical if $\omega(\gamma)$ exists for $A = E$ and is equal to $N(0,\sigma)$.

From Lemma 6.69 we obtain the following, which does not depend on $\phi$ being bicombable, but only weakly combable:

LEMMA 6.72. *Almost every $\gamma \in Y^i$ with respect to the measure $\widehat{\nu}$ is $E^i,\sigma^i$-typical.*

PROOF. The following proof was suggested by Shigenori Matsumoto.

We fix the notation below for the course of the Lemma (the reader should be warned that it is slightly incompatible with notation used elsewhere; this is done to avoid a proliferation of subscripts). Let $C_i$ be a component of $\Gamma$ with Perron–Frobenius root $\xi(C_i) = \lambda$. Let $Y_i$ be the set of infinite paths in $\Gamma$ 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 $\mu$ on $X$. The measure $\widehat{\mu}_i$ is determined by a stationary measure $\mu_i$ on $C_i$ and the transition matrix $N(i)$, the restriction of the measures $\mu$ and the matrix $N$ defined in § 6.5.3. The measure $\mu_i$ is stationary in the sense that $\mu_i^T N(i) = \mu_i^T$, so $\widehat{\mu}_i$ is shift invariant. Since $C_i$ is recurrent, $\mu_i^T$ is the only eigenvector of $N(i)$ with eigenvalue 1, so $\mu_i$ is extremal in the space of stationary measures. Therefore by the random ergodic theorem (see e.g. [**168**], Ch. 10) the measure $\widehat{\mu}_i$ on $X_i$ is ergodic.

Now, there is a subset $X_i^*$ of $X_i$ of full measure such that for all $\gamma \in X_i^*$,

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

in the weak* topology, where $S$ denotes shift map, and $\delta$ is a Dirac mass. On the other hand, on $Y_i$ there is a measure $\widehat{\nu}_i$ which is the restriction of $\widehat{\nu}$. Define $q : Y_i \to X_i$ by

$$q(\gamma) = S^{n(\gamma)}(\gamma)$$

where $n : Y_i \to \mathbb{N}$ satisfies the following condition. Let $\pi : X_i \to C_i$ take each infinite walk to its initial vertex. Choose $n$ so that $\pi \circ q : Y_i \to C_i$ sends the measure $\widehat{\nu}_i$ on $Y_i$ to a measure $\mu_q$ on $C_i$ of full support. The measure $q_*\widehat{\nu}_i$ on $X_i$ is obtained from an initial measure $\mu_q$ and the transition matrix $N(i)$ as in § 6.5.3; it follows that the measures $q_*\widehat{\nu}_i$ and $\mu_i$ are equivalent (i.e. each is absolutely continuous with respect to the other).

It follows that $Y_i^* := q^{-1}(X_i^*)$ has full measure with respect to $\widehat{\nu}_i$, and if $\gamma \in Y_i^*$, then

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

This shows that the geodesic ray in $G$ associated to any $\gamma \in Y_i^*$ is $E^i, \sigma^i$-typical, and the lemma is proved. □

On the other hand, the following Lemma uses bicombability in an essential way:

LEMMA 6.73. *Let $\gamma$ be an $E, \sigma$-typical geodesic ray in $G$. If $\phi$ is combable and if $\gamma'$ is a geodesic ray with the same endpoint at $\gamma$, then $\gamma'$ is also $E, \sigma$-typical. If $\phi$ is bicombable then for any $g \in G$, the translate $g\gamma$ is $E, \sigma$-typical.*

PROOF. Let $\gamma$ and $\gamma'$ have the same endpoint. Then there is a constant $C$ such that $d_L(\gamma_i, \gamma_i') \le C$ and therefore $|\phi(\gamma_i) - \phi(\gamma_i')| \le K$ for some $K$ independent of $i$. This shows that $\gamma'$ is $E, \sigma$-typical if $\gamma$ is. Similarly, if $g \in G$ then $d_R(g\gamma_i, \gamma_i) \le C$ and therefore $|\phi(g\gamma_i) - \phi(\gamma_i)| \le K$ for some $K$ independent of $i$. □

We now come to the crucial point. For each $i$, let $\partial^i G$ denote the image of the typical elements in $Y^i$ under the endpoint map $Y \to \partial G$. Note that $\widehat{\nu}(\partial^i G)$ is strictly positive for each $i$. By Theorem 6.63, for any $i, j$ there is some $g \in G$ with $\widehat{\nu}(g\partial^i G \cap \partial^j G) > 0$. It follows that there is a typical $\gamma \in Y^i$ and a typical $\gamma' \in Y^j$ such that (identifying elements of $Y$ and geodesic rays in $G$ starting at id) the translate $g\gamma$ and $\gamma'$ are asymptotic to the same endpoint in $\partial G$. Since $\phi$ is bicombable, by Lemma 6.73, $\gamma$ and $\gamma'$ are both $E^i, \sigma^i$-typical and $E^j, \sigma^j$-typical. It follows that $E^i = E^j$ and $\sigma^i = \sigma^j$. Together with Lemma 6.69, this proves the Central Limit Theorem:

THEOREM 6.74 (Central Limit Theorem; Calegari–Fujiwara, [**50**]). *Let $\phi$ be a bicombable function on a word-hyperbolic group $G$. Let $\overline{\phi}_n$ be the value of $\phi$ 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 average of $d\phi$ on $\Gamma$ with respect to the stationary measure.*

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

COROLLARY 6.75. *Let $\phi$ be a bicombable function on a word-hyperbolic group $G$. Then there is a constant $E$ such that for any $\epsilon > 0$ there is a $K$ and an $N$ so that if $G_n$ denotes the set of elements of length $n \geq N$, 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 symmetric generating sets for $G$. Word length in the $S_2$ metric is a bicombable function with respect to a combing $L_1$ for the $S_1$ generating set. Hence:

COROLLARY 6.76. *Let $S_1$ and $S_2$ be finite generating sets for $G$. There is a constant $\lambda_{1,2}$ such that for any $\epsilon > 0$, there is a $K$ and an $N$ so that if $G_n$ denotes the set of elements of length $n \geq N$ 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 equality*

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

REMARK 6.77. In Corollary 6.76 it is important to note that though a typical geodesic word of length $n$ in the $S_1$ metric is represented by a geodesic word of length $n\cdot\lambda_{1,2}$ in the $S_2$ metric, with error of order $\sqrt{n}$, the resulting set of geodesic words in the $S_2$ metric are not themselves typical. Thus $\lambda_{1,2}\lambda_{2,1} > 1$ in general. We give an example to illustrate this phenomenon in § 6.5.5.

If $\phi$ is a quasimorphism, then $|\phi(g)+\phi(g^{-1})| \leq$ const. so if $S$ is symmetric, then necessarily $E$ as above is equal to 0. Hence:

COROLLARY 6.78. *Let $\phi$ be a bicombable quasimorphism on a word-hyperbolic group $G$. Let $\overline{\phi}_n$ be the value of $\phi$ 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 $\sigma \geq 0$.*

**6.5.5. An example.** Let $F$ denote the free group on two generators $a, b$. Let $S_1$ denote the symmetric generating set $S_1 = \langle a, b, a^{-1}, b^{-1}\rangle$ and $S_2$ the symmetric generating set $S_2 = \langle a, b, c, a^{-1}, b^{-1}, c^{-1}\rangle$ where $c = ab$ (and therefore $c^{-1} = b^{-1}a^{-1}$). We compare word length in the $S_1$ and the $S_2$ metrics.

One can verify that a word in the $S_2$ generating set is a geodesic if and only if it is reduced, and contains no subwords of the form $a^{-1}c, cb^{-1}, c^{-1}a, c^{-1}b$, and moreover that geodesic representatives are unique. The language of all geodesics in the $S_2$ generating set is therefore a combing.

One can build a digraph $\Gamma$ which parameterizes the language of geodesics in $S_2$ as follows. There are seven vertices, one initial vertex and six other vertices labeled by the elements of $S_2$. There is an outgoing edge from the initial vertex to each other vertex, and one directed edge from $x$ to $y$ for each other vertex if and only if $xy$ is not one of the four "excluded" words above. See Figure 6.7. The vertices have been labeled $a, b, c, A, B, C$ and labels have been left off the edges for clarity.

Let $\Gamma'$ be obtained from $\Gamma$ by removing the initial vertex. The adjacency matrix of $\Gamma'$ is

$$M = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

which is Perron–Frobenius with biggest real eigenvalue 4, and $\mathbf{1}^T, \mathbf{1}$ as left and right eigenvectors. It follows that the stationary measure is just equal to the ordinary uniform measure. Note that there are $6 \times 4^{n-1}$ words of length $n$ in the $S_2$ metric, and $4 \times 3^{n-1}$ words of length $n$ in the $S_1$ metric.

Let $\phi_{S_i}$ denote the bicombable function which computes word length in the $S_i$ metric. There are discrete derivatives $d\phi_{S_1}, d\phi_{S_2}$ from the vertices of $\Gamma'$ to 1. Here $d\phi_{S_2}$ is just the constant function $\Gamma' \to 1$, whereas $d\phi_{S_1}$ takes the value 1 on the vertices labeled $a, b, A, B$ and 2 on the vertices labeled $c, C$. It follows that a random word of length $n$ in the $S_2$ metric has length $4n/3$ in the $S_1$ metric, with error of order $\sqrt{n}$.

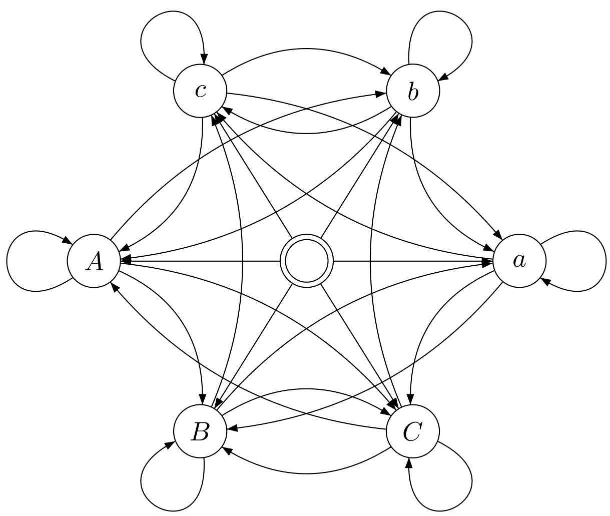


FIGURE 6.7. A digraph parameterizing geodesics in the $S_2$ metric

On the other hand, $d\phi_{S_1}$ and $d\phi_{S_2}$ exist as functions from the vertices of $\Gamma_1'$ to 1 where $\Gamma_1'$ is the digraph in Figure 6.4. In this case, $d\phi_{S_1}$ is the constant function $\Gamma_1' \to 1$ and $d\phi_{S_2}$ is the function which takes the value 0 on the vertices labeled $ab$ and $b^{-1}a^{-1}$, and 1 on all other vertices. It follows that a random word of length $n$ in the $S_1$ metric has length $5n/6$ in the $S_2$ metric, with error of order $\sqrt{n}$. Hence $\lambda_{1,2}\lambda_{2,1} = 5/6 \times 4/3 = 10/9$. with notation as in Corollary 6.76.

In general, if the growth rate in the $S_i$ metric is $\lambda_i^n$ for $i = 1, 2$ then there is an inequality $\lambda_{i,j} \geq \log \lambda_i / \log \lambda_j$, by counting. In this case, we get the two (easily verified) inequalities

$$0.83333\cdots = \frac{5}{6} \geq \frac{\log 3}{\log 4} = 0.79248\cdots$$

and

$$1.33333\cdots = \frac{4}{3} \geq \frac{\log 4}{\log 3} = 1.26186\cdots$$

REMARK 6.79. The numbers $\lambda_{1,2}$ where $S_1$ and $S_2$ are a symmetric basis for a free group $F_k$, are studied in [**119**], where it is shown that they are always rational, and satisfy $2k\lambda \in \mathbb{Z}[1/(2k-1)]$.