A CENTRAL LIMIT THEOREM FOR RANDOM CLOSED GEODESICS: PROOF OF THE CHAS–LI–MASKIT CONJECTURE

ILYA GEKHTMAN, SAMUEL J. TAYLOR, AND GIULIO TIOZZO

Abstract. We prove a central limit theorem for the length of closed geodesics in any compact orientable hyperbolic surface. In the special case of a hyperbolic pair of pants, this settles a conjecture of Chas–Li–Maskit.

1. Introduction

Let $\Sigma$ be a compact orientable hyperbolic surface whose boundary, if any, is geodesic, and let $G$ denote its fundamental group. By a standard generating set $S$ for $G$ we mean the following: when $G$ is free (i.e. when $\partial \Sigma \neq \emptyset$) $S$ is a free basis for $G$. Otherwise, $\Sigma$ is a closed orientable surface of genus $g \geq 2$ and $S$ is the generating set used in the usual presentation $G = \langle a_1, \ldots, a_g, b_1, \ldots b_g : \prod [a_i, b_i] = 1 \rangle$.

Now fix a standard generating set $S$ of $G$, and let $|g|$ be the word length of $g$ with respect to $S$. For each $g \in G$, let $[g]$ denote its conjugacy class, and for any conjugacy class $\gamma = [g]$ define its conjugacy length $\|\gamma\| = \|g\| := \min_{[g]=\gamma} |g|$ to be the minimum word length over all elements representing $\gamma$.

Any conjugacy class $\gamma$ is represented by a closed geodesic in $\Sigma$, and let $\tau(\gamma)$ denote the length of this geodesic in the hyperbolic metric. Let $\mu_n$ denote the uniform distribution on the set $\mathcal{F}_n$ of conjugacy classes of length $n$. The goal of this note is to prove the following central limit theorem:

Theorem 1. There exist constants $L > 0, \sigma > 0$ such that for any $a, b \in \mathbb{R}$ with $a < b$ we have

$$
\mu_n \left( \gamma : \frac{\tau(\gamma) - nL}{\sigma \sqrt{n}} \in [a, b] \right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} \, dx
$$

as $n \rightarrow \infty$.

Motivated by experimental evidence, Chas–Li–Maskit [7] conjectured that the conclusion of Theorem 1 holds for a hyperbolic pair of pants. This followed an earlier central limit theorem by Chas–Lalley [6] for the distribution of self-intersection numbers of random geodesics.

The proof of Theorem 1 uses the central limit theorem for Hölder continuous observables on a mixing Markov chain following Ruelle [20] and Bowen [2] (see also Pollicott–Sharp [17] and Calegari [4]), combined with estimates on Gromov products by the authors [10]. The Markov chain which encodes closed geodesics on a surface is provided by Series [21, 22, 23] (see also Wroten [25]).

Date: June 12, 2019.

Gekhtman is partially supported by NSF grant DMS-1401875.

Taylor is partially supported by NSF grant DMS-1744551.

Tiozzo is partially supported by NSERC and the Alfred P. Sloan Foundation.
We note that P. Park has recently written up a related result where the uniform distribution on conjugacy classes is replaced by the $n^{th}$ step distribution of a simple random walk on $G$ [16]. Let us note that counting for the simple random walk and counting with respect to balls in the Cayley graph are in general different, and many authors addressed the question of how they are related on various groups.

Acknowledgments. We thank the referee for useful suggestions that improved our exposition.

2. Preliminaries

The geometric setup. Since our argument will use tools from coarse geometry, we begin with some basic definitions; additional background can be found in [4].

Let $(X,d)$ be a $\delta$-hyperbolic geodesic metric space, for some $\delta > 0$. Recall this means that between any two points $x,y \in X$ there is some geodesic segment $[x,y]$ in $X$, and for any geodesic triangle $[x,y],[y,z],[z,x]$ in $X$ with $x,y,z \in X$, one has the inclusion $[x,y] \subset N_\delta ([y,z] \cup [z,x])$, where $N_\delta$ denotes the $\delta$-neighborhood in $X$. In this paper, $X$ will usually be either the hyperbolic plane $\mathbb{H}^2$ or the Cayley graph of a free or surface group with respect to a fixed generating set. These are standard examples of $\delta$-hyperbolic spaces (for different $\delta$).

For $x,y,z \in X$, the Gromov product $(x,y)_z$ is defined to be

$$(x,y)_z = \frac{1}{2} (d(z,x) + d(z,y) - d(x,y)).$$

We now specialize to the case of interest, where $G = \pi_1(\Sigma)$ for some hyperbolic surface $\Sigma$ as in the introduction. Throughout we identify the universal cover $\tilde{\Sigma}$ with a convex subspace of the hyperbolic plane $\mathbb{H}^2$ and consider $G$ as a discrete group of isometries of $\mathbb{H}^2$. When $\Sigma$ is closed, $G$ is a cocompact Fuchsian group. Otherwise, $\partial \Sigma \neq \emptyset$ and $G$ is the free group $F_N$ for some $N \geq 2$. In this case, we have that $\Sigma$ is the convex core of $\mathbb{H}^2/G$, where $G$ acts on $\mathbb{H}^2$ as a Schottky group.

Fix a base point $z \in \mathbb{H}^2$. Then for $\gamma = [g]$, $\tau(\gamma) = \tau(g)$ equals the (stable) translation length of $g$ on $\mathbb{H}^2$. Hence, one has the formula (see e.g. [15, Proposition 5.8])

(1) $$\tau(g) = d(gz,z) - 2(gz,g^{-1}z)_z + O(\delta)$$

where $(x,y)_z$ is the Gromov product in $\mathbb{H}^2$. Here $A = B + O(\delta)$ means that there is a constant $C$, which depends only on the hyperbolicity constant $\delta$ of $\mathbb{H}^2$, such that $|A - B| \leq C$. (In fact, a standard computation shows that one can take $\delta = \log(1 + \sqrt{2})$, but we will not need this fact.)

Some basic probability. We begin by recording a few basic lemmas that will be needed for our arguments.

**Lemma 2.** Let $(A_n)$ be any sequence of measurable sets in a probability space, $(\mathbb{P}_n)$ a sequence of probability measures, and let $(B_n)$ be a sequence such that

$$\lim_{n \to \infty} \mathbb{P}_n(B_n) = 1.$$ 

Then

$$\limsup_n |\mathbb{P}_n(A_n) - \mathbb{P}_n(A_n \cap B_n)| = 0.$$
Proof. By elementary set theory,
\[ P_n(A_n) = P_n(A_n \cap B_n) + P_n(A_n \setminus B_n) \leq P_n(A_n \cap B_n) + P_n(B_n^c) \]
which yields the claim. \[ \square \]

**Lemma 3.** Let \((P_n)\) be a sequence of probability measures on a Borel space \(X\), and let \(F, G : X \to \mathbb{R}\) be two measurable functions. Suppose that \(P_n(|G(x)| \geq \epsilon) \to 0\) for any \(\epsilon > 0\), and let \((c_n)\) be a sequence of positive real numbers with \(\lim_{n \to \infty} c_n = 1\). Suppose that there exists a continuous function \(\rho : \mathbb{R} \to \mathbb{R}^+\) such that
\[ \lim_{n \to \infty} P_n(F(x) \in [a, b]) = \int_a^b \rho(x) \, dx. \]
Then
\[ \lim_{n \to \infty} P_n\left(\frac{F(x) + G(x)}{c_n} \in [a, b]\right) = \int_a^b \rho(x) \, dx. \]

**Proof.** Fix \(\epsilon > 0\), and denote \(\Phi_{a,b} := \int_a^b \rho(x) \, dx\). If \(n\) is sufficiently large, then by setting \(Y_n := \frac{F+G}{c_n}\) we have
\[ \{F(x) \in [a + \epsilon, b - \epsilon] \text{ and } |G(x)| \leq \epsilon/2\} \subseteq \{Y_n \in [a, b]\} \]
and using Lemma 2
\[ \liminf P_n(Y_n \in [a, b]) \geq \liminf P_n(F \in [a + \epsilon, b - \epsilon] \text{ and } |G| \leq \epsilon/2) = \liminf P_n(F \in [a + \epsilon, b - \epsilon]) = \Phi_{a+\epsilon,b-\epsilon}. \]
On the other hand
\[ \{Y_n \in [a, b] \text{ and } |G| \leq \epsilon/2\} \subseteq \{F \in [a - \epsilon, b + \epsilon]\} \]
hence
\[ \limsup P_n(Y_n \in [a, b]) = \limsup P_n(Y_n \in [a, b] \text{ and } |G| \leq \epsilon/2) \leq \limsup P_n(F \in [a - \epsilon, b + \epsilon]) = \Phi_{a-\epsilon,b+\epsilon} \]
and taking \(\epsilon \to 0\) completes the proof. \[ \square \]

For the following lemma, recall that the total variation of a signed measure \(\mu\) on a measure space \((X, \mathcal{A})\) is defined as \(\|\mu\|_{TV} := \sup_{A \in \mathcal{A}} |\mu(A)|\), where the supremum is taken over all measurable subsets \(A \subseteq X\).

**Lemma 4.** If \(\lambda, \nu\) are purely atomic probability measures on a set \(X\) and \(\lambda\) is absolutely continuous with respect to \(\nu\), then
\[ \|\lambda - \nu\|_{TV} \leq \left\| \frac{d\lambda}{d\nu} - 1 \right\|_{\infty}, \]
where \(\|\cdot\|_{TV}\) denotes the total variation of a measure.

**Proof.** Since the measures are atomic,
\[ \frac{d\lambda}{d\nu}(x) = \frac{\lambda(x)}{\nu(x)} \quad \text{for any } x \in X. \]
Then
\[ \|\lambda - \nu\|_{TV} = \sup_{A \subseteq X} |\lambda(A) - \nu(A)| \leq \sup_{A} \sum_{x \in A} |\lambda(x) - \nu(x)| \leq \left\| \frac{d\lambda}{d\nu} - 1 \right\|_{\infty}. \]
paths in $\Gamma$ to $G$ map to geodesics in $G$ in item (1) of Lemma 5. We further remark that Lemma 5 implies that directed paths in $\Gamma$ also yield this result. The much easier case of free groups is briefly explained below.

Coding for closed geodesics. First of all, we use that conjugacy classes in free and surface groups can be encoded by a finite graph.

**Lemma 5** ([21, 19]). Let $G = \pi_1(\Sigma)$ where $\Sigma$ is an orientable hyperbolic surface of finite type. Let $S$ be a standard generating set for $G$. Then there exists an oriented graph $\Gamma$ whose edges are labeled by elements of $S \cup S^{-1}$ and such that:

1. cycles in $\Gamma$ of length $n$ are in bijection with conjugacy classes $C(G)$ of length $n$ in the group, except for finitely many exceptions;
2. a conjugacy class is primitive if and only if the corresponding cycle in the graph is primitive, i.e. not the power of a shorter cycle;
3. the adjacency matrix $M$ for this graph is aperiodic.

For closed surface groups, Lemma 5 is precisely ([19], Lemma 1.1), and we direct the reader there for a proof which assembles work of Series [21, 22, 23]. Very briefly, this coding goes back to Bowen–Series [3] who use the action $G \acts \partial \mathbb{H}^2 = \mathbb{S}^1$ to build a (Markov) map $f : \mathbb{S}^1 \to \mathbb{S}^1$ having the following property: there is a partition of $\mathbb{S}^1$ into a finite union of intervals $\{J_i\}_{i=1}^r$, disjoint except at their endpoints, such that for each $i$, $f(J_i)$ is a union of intervals in the partition. Moreover, for each $i$ there is a generator $s_i$ in $S \cup S^{-1}$ such that $f$ and $s_i^{-1}$ have the same restriction to $J_i$. The graph $\Gamma$ is then the oriented graph whose vertices are the intervals $\{J_i\}$ with a directed edge labeled $s_i$ from $J_i$ to $J_j$ if $J_j \subset f(J_i)$. Another coding for closed surfaces is given by Wrotten [25] and also yields this result. The much easier case of free groups is briefly explained below.

Recall that a matrix $M$ is aperiodic if there exists an integer $k \geq 1$ such that all entries of $M^k$ are positive. By the Perron-Frobenius theorem, the matrix $M$ has a unique, simple eigenvalue $\lambda > 0$ of maximum modulus. Also, if $\nu$ denotes the map which reads the labels off of oriented edges of $\Gamma$, then $\nu$ extends to the evaluation map from directed paths in $\Gamma$ to $G$. In details, if a directed path $p$ is a concatenation of oriented edges $e_1, \ldots, e_n$, then $\nu(p) = \nu(e_1) \cdots \nu(e_n)$ in $G$. It is this map that induces the bijection in item (1) of Lemma 5. We further remark that Lemma 5 implies that directed paths in $\Gamma$ map to geodesics in $G$; that is, if $p$ is a directed path of length $n$ in $\Gamma$, then $|\nu(p)| = n$ with respect to the generating set $S$.

We note that for free groups, the construction of the graph $\Gamma$ is immediate. Let $G = F_N$ and fix a basis $\{a_1, \ldots, a_N\}$ of $F_N$. Then the graph $\Gamma$ has $2N$ vertices, labelled $a_i^\epsilon$ with $i = 1, \ldots, N$, $\epsilon = \pm$. For each vertex $v = a_i^\epsilon$, there exists an edge labelled $a_i^\eta$ to the vertex $a_j^\eta$ unless $i = j$ and $\epsilon = -\eta$. In this case, nontrivial cyclically reduced words (i.e. words that do not end with the inverse of their first letter) are in bijection with oriented

\[
\leq \sup_A \sum_{x \in A} \left| \frac{\lambda(x)}{\nu(x)} - 1 \right| \nu(x) = \sup_A \left( \left\| \frac{d\lambda}{d\nu} - 1 \right\| \nu(A) \right) \leq \left\| \frac{d\lambda}{d\nu} - 1 \right\|_\infty ,
\]

completing the proof. \qed

3. The central limit theorem for displacement

The proof of Theorem 1 uses equation (1): basically, one proves a CLT for the displacement function $d(z, gz)$, and then shows that the contribution of the term $2(gz, g^{-1}z)$ tends to zero. These two facts will suffice by Lemma 3. We will start by establishing the CLT for displacement.
based closed paths since such a word can only be read as a closed path by starting at
the vertex corresponding to the label on its last letter. As two cyclically reduced words
represent conjugate elements if and only if they differ by cyclic permutation, which
corresponds to changing the basepoint of the loop, this establishes items (1) and (2). (In
this case, the only exception in item (1) is for the trivial conjugacy class, which can be
read as the trivial cycle based at each vertex.) Item (3) is clear from the construction.

Let $\lambda_n$ be the uniform distribution on the set $C_n$ of (based) closed paths of length $n$ in
$\Gamma$. A short counting argument shows the following, which is a variation of ([6], Lemma
4.1):

**Lemma 6.** Let $p_n : C_n \to C_n$ be the map from closed paths to conjugacy classes induced
by the evaluation map. Then

$$
\|(p_n)_* \lambda_n - \mu_n\|_{TV} \to 0 \quad \text{as } n \to \infty,
$$

where $\| \cdot \|_{TV}$ denotes the total variation of a measure.

**Proof.** Let $\Gamma_n$ denote the set of primitive cycles of length $n$ (without remembering the
basepoint). Note that $Tr M^n = \lambda^n + \sum_{i=1}^k \lambda_i^n$, where $|\lambda_i| < \lambda$. Hence $\lambda^n \leq Tr M^n \leq c\lambda^n$ where $c$ only depends on $M$, and $\lim_{n \to \infty} \frac{Tr M^n}{\lambda^n} = 1$. Moreover, by definition

$$
\sum_{d|n} d(#\Gamma_d) = Tr M^n, \text{ so } n(#\Gamma_n) \leq n M^n \leq c\lambda^n.
$$

Hence, the set of non-primitive closed paths of length $n$ has cardinality

$$
Tr M^n - n(#\Gamma_n) = \sum_{d|n, d \neq n} d(#\Gamma_d) \leq \frac{cn}{2} \lambda^{n/2}
$$

and so

$$
\lim_{n \to \infty} \frac{n(#\Gamma_n)}{\lambda^n} = 1.
$$

Then for any set $A \subseteq C_n$ and any $n \geq 1$ sufficiently large (using the bijection provided
by Lemma 5 with $n$ large enough to avoid the exceptions)

$$
\mu_n(A) = \frac{\sum_{d|n} d(#(A \cap \Gamma_d))}{\sum_{d|n} d(#\Gamma_d)} = \frac{#(A \cap \Gamma_n) + O(n\lambda^{n/2})}{(#\Gamma_n) + O(n\lambda^{n/2})}
$$

and, since the map $p_n$ is exactly $d$-to-1 on the preimage of $\Gamma_d$,

$$
\lambda_n(p_n^{-1}(A)) = \frac{\sum_{d|n} d(#(A \cap \Gamma_d))}{\sum_{d|n} d(#\Gamma_d)} = \frac{n\#(A \cap \Gamma_n) + O(n\lambda^{n/2})}{n(#\Gamma_n) + O(n\lambda^{n/2})}.
$$

Hence

$$
|\mu_n(A) - \lambda_n(p_n^{-1}(A))| = O\left(\frac{\lambda^{n/2}}{\#\Gamma_n}\right) \to 0,
$$

as $n \to \infty$. This completes the proof. \qed
Markov chains. Let $\Gamma$ be a directed graph with vertex set $V = V(\Gamma)$, edge set $E = E(\Gamma)$, and aperiodic transition matrix $M$. For $n \geq 0$, we let $\Omega^n$ denote the set of paths of length $n$ in $\Gamma$ starting at any vertex, and $\Omega^* = \bigcup_{n \geq 1} \Omega^n$ the set of all paths of finite length. Also denote by $\Omega$ the set of all one-sided infinite directed paths starting at vertices of $\Gamma$. If we wish to focus on the subset of paths that start at the vertex $v \in V$, then we use $v$ as a subscript as in $\Omega_v$ or $\Omega^*_v$. We will associate to $\Gamma$ a probability on each edge and a probability on each vertex so that the corresponding measure on the space of infinite paths in $\Gamma$ is shift-invariant. (The reader is directed to [24, Section 8] for additional details of this standard construction.) Recall that if $x = (x_i)_{i \geq 0} \in \Omega$ is a directed path in $\Gamma$, then the image of $x$ under the shift $T: \Omega \to \Omega$ is $Tx = (x_i)_{i \geq 1}$. The assumption that $M$ is aperiodic translates to the fact that the shift is topologically mixing. For $x \in \Omega^n$, we continue to use the notation $T^ix$ to denote the image of $x$ under the shift for $i \leq n$.

Recall that $\lambda > 1$ is the unique (simple) eigenvalue of $M$ of maximum modulus. Fix a right eigenvector $v \equiv (\lambda)v$ for $M$ of eigenvalue $\lambda$ and a left eigenvector $u$ of the same eigenvalue, normalized so that $u^T v = 1$. Let us now define the measure $(\pi_i)_{i \in V}$ on the set of vertices of $\Gamma$ where $\pi_i = u_i v_i$, and for each edge in $\Gamma$ from $i$ to $j$ let us define the probability

$$q_{ij} = \frac{m_{ij}v_j}{\lambda v_i}.$$  

This defines a Markov chain on $\Gamma$, where $\pi_i$ is the probability of starting at vertex $v_i$, and $q_{ij}$ is the probability of going from $v_i$ to $v_j$.

This construction defines a shift invariant measure $\nu$ (the so-called Parry measure) on the set of infinite paths $\Omega$. Namely, let $C(i_0, i_1, \ldots, i_k)$ denote the set of infinite paths which start with the path $v_{i_0} \to v_{i_1} \to \cdots \to v_{i_k}$. This is a cylinder set of $\Omega$ and we set its measure to be

$$\nu(C(i_0, i_1, \ldots, i_k)) = \pi_{i_0} q_{i_0 i_1} q_{i_1 i_2} \cdots q_{i_{k-1} i_k},$$

and this determines a shift invariant measure $\nu$ on $\Omega$. (In fact, this defines the measure of maximal entropy for the shift $T: \Omega \to \Omega$ [24, Theorem 8.10].) Now for each $n$, let $\nu_n$ be the pushforward of $\nu$ with respect to the map $\Omega \to \Omega^n$ which takes an infinite path to its prefix of length $n$. The measure $\nu_n$ is the distribution of the $n$th step of the Markov chain whose initial distribution is $(\pi_i)_{i \in V}$, and $\nu_n$ is supported on the set of paths of length $n$.

As before $\text{ev} : E(\Gamma) \to G$ is the evaluation map which associates to each edge its label in $S \cup S^{-1} \subset G$. By concatenation, the map extends to a map $\text{ev} : \Omega^* \to G$ from the set of all finite paths to $G$. Hence, if $x \in \Omega^n$ is a path of length $n$ given as a sequence of vertices $(x_i)_{i=0}^n$ associated to the edge path $e_1, \ldots, e_n$, then $\text{ev}(x) = \text{ev}(e_1) \cdots \text{ev}(e_n)$ in $G$. In particular, if $x$ is a single vertex (i.e. a path of length 0) then $\text{ev}(x) = 1$.

3.1. The Central Limit Theorem for Hölder observables. Here we briefly recall the classical CLT from Thermodynamic Formalism as we will need it. We closely follow Bowen [2, Chapter 1].

Let $T : \Omega \to \Omega$ be a topologically mixing shift of finite type and $\nu$ a shift-invariant Gibbs measure on $\Omega$. (For us, $\nu$ will always be the measure of maximal entropy on the Markov chain $\Omega$, defined as above.) We have a metric $d_2$ on the space $\Omega \times \Omega$ of all directed paths defined by $d_2(x, y) = 2^{-K}$ where $K$ is the largest nonnegative integer
such that $x_i = y_i$ for $i < K$. We note that with this metric, $\Omega$ is the set of limit points of the discrete set $\Omega^*$. Topologically, $\Omega$ is a Cantor set.

A function $f: \Omega \to \mathbb{R}$ is Hölder continuous if for all $x, y \in \Omega$, $|f(x) - f(y)| \leq C_d(x, y)\epsilon$ for some $C, \epsilon > 0$.

The following result is a combination of Theorem 1.27 of [2] and the remark that follows it. It states that a central limit theorem holds with positive variance as long as the Livšic cohomological equation [14] has no solutions. We will use the notation $N_{a,b} := \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$.

**Theorem 7** ([2], Theorem 1.27). Suppose that $f: \Omega \to \mathbb{R}$ is Hölder and that there does not exist a Hölder function $u: \Omega \to \mathbb{R}$ such that $f = u - u \circ T + \int f dv$. Then there is a constant $\sigma > 0$ such that for any $a < b$,

$$
\nu \left( x \in \Omega : \frac{\sum_{i=0}^{n-1} f(T^i x) - n \int f dv}{\sigma \sqrt{n}} \in [a, b] \right) \to N_{a,b}.
$$

3.2. Displacement and the CLT. The main result of this section is the following central limit theorem for displacement along the Markov chain. First, fix a basepoint $z \in \mathbb{H}^2$.

**Theorem 8.** There exist constants $L > 0$ and $\sigma > 0$ such that for any $a, b \in \mathbb{R}$ with $a < b$ we have

$$
\nu_n \left( x \in \Omega^n : \frac{d(\text{ev}(x)z, z) - nL}{\sigma \sqrt{n}} \in [a, b] \right) \to N_{a,b}.
$$

The proof of Theorem 8 requires the following setup, which approximately follows the discussion in Calegari [4, Section 3.7]. For $g \in G$ and $s \in S \cup S^{-1}$, define $D_s F(g) = d(z, gz) - d(z, sgz)$. Note that by the triangle inequality

$$
|D_s F(g)| = |d(z, gz) - d(s^{-1} z, gz)| \leq d(z, s^{-1} z) \leq \max_{s \in S \cup S^{-1}} d(z, sz).
$$

For a finite path $x \in \Omega^*$, let $DF(x) = D_{s^{-1}} F(\text{ev}(x))$, where $s$ labels the first edge of $x$. This defines a function $DF: \Omega^* \to \mathbb{R}$ on the set of all finite paths such that

$$
DF(x) = d(z, \text{ev}(x)z) - d(z, \text{ev}(Tx)z).
$$

In particular, if $x \in \Omega^n$, then we note that

$$
\sum_{i=0}^{n-1} DF(T^i x) = d(z, \text{ev}(x)z).
$$

To apply the central limit theorem (Theorem 7), one needs to verify the Hölder continuity property of the observable, and so we use the following proposition of Pollicott and Sharp. For the statement, if $g, h \in G$, then $(g, h)$ will denote their Gromov product based at the identity $1 \in G$. That is,

$$
(g, h) := (g, h)_1 = \frac{1}{2} \left( |g| + |h| - |g^{-1} h| \right).
$$

We note that this quantity is always positive by the triangle inequality.
Lemma 9 ([17], Proposition 1; [18], Lemma 1 and Proposition 3). There exist constants $C > 0$ and $\alpha > 1$ such that for any $h, g \in G$, any $s \in S \cup S^{-1}$

$$|D_s F(g) - D_s F(h)| \leq C \alpha^{-\langle g, h \rangle}.$$  

(5)

Since the evaluation map $ev: \Omega^* \to G$ is geodesic, we claim that for any $x, y \in \Omega^*$

$$\alpha^{-\langle ev(x), ev(y) \rangle} \leq \alpha d_{\Omega}(x, y)^\eta$$

where $\eta = \log_2(\alpha)$. Indeed, suppose $d_{\Omega}(x, y) = 2^{-K}$ so that we can write $x = px'$ and $y = py'$ where $p$ is the (possibly empty) length $k = K - 1$ common prefix of $x$ and $y$ in $\Omega^*$. (Recall that $K$ is the number of initial vertices in common.) Then directly from the definition of the Gromov product and the fact that $ev: \Omega^* \to G$ maps paths of length $n$ to elements of word length $n$, we have $\langle ev(x), ev(y) \rangle = k + \langle ev(x'), ev(y') \rangle \geq K - 1$. From this the claim is evident.

This observation plus Lemma 9 directly implies that $DF: \Omega^* \to \mathbb{R}$ is Hölder. For this, let $x, y \in \Omega^*$. First, note that $|DF(x) - DF(y)| \leq 2 \max_{s \in S \cup S^{-1}} d(z, sz)$ by (3), and this suffices whenever $x$ and $y$ do not share an initial edge, since in this case $d_{\Omega}(x, y) \geq 1/2$. Otherwise, $x$ and $y$ have a common initial edge with label $s \in S \cup S^{-1}$ and Lemma 9 gives

$$|DF(x) - DF(y)| = |D_{s^{-1}} F(ev(x)) - D_{s^{-1}} F(ev(y))|$$

$$\leq C \alpha^{-\langle ev(x), ev(y) \rangle}$$

$$\leq C \alpha d_{\Omega}(x, y)^\eta,$$

as required.

Hence, the function $DF$ on $\Omega^*$ is Hölder continuous and therefore has a unique continuous extension to the Hölder function $DF: \Omega \to \mathbb{R}$ on the space of all infinite paths $\Omega$.

For $x \in \Omega$, define

$$F_n(x) = DF(x) + DF(Tx) + \ldots + DF(T^{n-1}x).$$

If we let $x^n \in \Omega^n$ denote the prefix of $x$ of length $n$, we have that

$$F_n(x^n) = d(z, ev(x^n)z),$$

by (4). Whereas for an infinite path $x \in \Omega$, we have

$$F_n(x) = \lim_{k \to \infty} \left( d(z, ev(x^k)z) - d(z, ev(T^n x^k)z) \right).$$

However, the following lemma bounds how far $F_n(x)$ can be from the displacement $F_n(x^n) = d(z, ev(x^n)z)$.

Lemma 10. The difference $|F_n(x) - F_n(x^n)|$ is uniformly bounded, independent of $x \in \Omega$ and $n \geq 1$. 
Proof. Write \( x = y_1 y_2 \ldots \), where the \( y_i \) are edges of \( \Gamma \), i.e. we represent \( x \) as an edge path in \( \Gamma \). Then \( x^k = y_1 y_2 \ldots y_k \). Also, let \( \text{ev}(y_i) = g_i \). Then

\[
|F_n(x^n) - F_n(x)| = \lim_{k \to \infty} \left( d(z, g_1 \ldots g_n z) + d(z, g_{n+1} \ldots g_k z) - d(z, g_1 \ldots g_k z) \right)
\]

\[
= \lim_{k \to \infty} \left( d(z, g_1 \ldots g_n z) + d(g_1 \ldots g_n z, g_1 \ldots g_k z) - d(z, g_1 \ldots g_k z) \right)
\]

\[
= 2 \lim_{k \to \infty} \left( z, \text{ev}(x^k z) \right)_{\text{ev}(x^n) z}.
\]

Since \( G \curvearrowright \mathbb{H}^2 \) is convex cocompact and the path \( i \to \text{ev}(x^i) = g_1 \ldots g_i \) is geodesic in \( G \), the path \( i \to \text{ev}(x^i)z \) is a uniform quasigeodesic. That is, \( i \to \text{ev}(x^i)z \) is a \((K,C)\)-quasigeodesic in \( \mathbb{H}^2 \) for \( K \geq 1 \) and \( C \geq 0 \) not depending on \( x \). This (together with the stability of quasigeodesics in the hyperbolic space \( \mathbb{H}^2 \)) immediately implies that the quantity \( (z, \text{ev}(x^k z) \text{ev}(x^n) z) \) is uniformly bounded for every \( k \geq n \). This completes the proof. \( \square \)

Finally, the following lemma is needed to establish positivity of \( \sigma \) in Theorem 8.

Lemma 11. There does not exist a function \( u : \Omega \to \mathbb{R} \) and \( L \in \mathbb{R} \) such that

\[
DF = u - u \circ T + L.
\]

Proof. Suppose not. Then

\[
\sum_{i=0}^{n-1} DF(T^i(x)) = u(x) - u(T^n(x)) + nL
\]

Thus, if \( x \in \Omega \) is a periodic point of period \( n \) for \( T \), then

\[
F_n(x) = \sum_{i=0}^{n-1} DF(T^i(x)) = nL.
\]

Next, a direct computation shows that if \( g = \text{ev}(x^n) \), then

\[
\tau(g) = \tau(\text{ev}(x^n)) = F_n(x) = nL.
\]

Indeed, as in [12, Lemma 4.2], since \( T^n x = x \), \( F_{nk}(x) = kF_n(x) \) and we see that \( F_{nk}(x) = F_{nk}(x^{nk}) + O(1) \) (Lemma 10) implies

\[
F_n(x) = \lim_{k \to \infty} \frac{1}{k} F_{nk}(x^{nk}) = \lim_{k \to \infty} \frac{1}{k} d(z, g^k z),
\]

which, by definition, equals the translation length \( \tau(g) \).

Hence, we conclude using Lemma 5 that \( \tau([g]) = L||g|| \) for all but finitely many conjugacy classes in \( G \). This, however, contradicts the fact that in every noncyclic subgroup of \( G \) there exist conjugacy classes whose geodesic representatives on \( \Sigma \) have incommensurable lengths (see, for example, [13]). \( \square \)

Proof of Theorem 8. Since \( DF \) is a Hölder continuous function on a mixing shift of finite type, Theorem 7 along with Lemma 11 give

\[
\nu \left( x : \frac{F_n(x) - nL}{\sigma \sqrt{n}} \in [a, b] \right) \to N_{a,b},
\]
where as before \( F_n(x) = DF(x) + DF(Tx) + \ldots + DF(T^{n-1}x) \). But by Lemma 10, the probability
\[
\nu \left( x : \frac{|F_n(x^n) - F_n(x)|}{\sqrt{n}} \geq \epsilon \right) \rightarrow 0,
\]
as \( n \rightarrow \infty \), for any \( \epsilon > 0 \). Hence, applying Lemma 3 gives the CLT for the displacement \( F_n(x^n) = d(x, ev(x^n)z) \) for \( x \in \Omega \) with respect to the measure \( \nu \). Since the distribution of \( x^n \in \Omega \) is \( \nu_n \), this completes the proof of the theorem.

4. CONVERGENCE TO THE COUNTING MEASURE FOR CLOSED PATHS

Next we use the Theorem 8 (which is about the Markov chain) to study the distribution of closed paths. Given a path \( x \) of length \( n \), let \( \tilde{x} \) denote the prefix of \( x \) of length \( n - \log n \).

Recall that \( \lambda_n \) is the uniform distribution on the set \( C_n \) of based closed paths of length \( n \) in \( \Gamma \). Since \( C_n \subset \Omega \), \( \lambda_n \) defines a measure on \( \Omega \) supported on \( C_n \). Let \( \lambda_{n,m} \) denote the distribution of the prefix of length \( n - m \) of a uniformly chosen closed path of length \( n \). Said differently, \( \lambda_{n,m} \) is the distribution on paths of length \( n - m \) obtained by pushing \( \lambda_n \) forward under the prefix map. In particular, if \( x \) has distribution \( \lambda_n \), then \( \tilde{x} \) has distribution \( \lambda_{n,\log n} \). We define \( \nu_{n,m} \) in the same way, using \( \nu_n \) in place of \( \lambda_n \).

**Proposition 12.** With notation as above, \( \lambda_{n,m} \) is absolutely continuous with respect to \( \nu_{n,m} \), and moreover
\[
\sup_{\gamma \in \Omega^{n-m}} \left| \frac{d\lambda_{n,m}(\gamma)}{d\nu_{n,m}(\gamma)} - 1 \right| \rightarrow 0
\]
as \( \min\{m, n\} \rightarrow \infty \).

**Proof.** Given a path \( \gamma = e_1 \cdots e_{n-m} \) with starting vertex \( v_i \) and end vertex \( v_j \) we have
\[
\lambda_{n,m}(\gamma) = \frac{\#\{\text{paths of length } m \text{ from } v_j \text{ to } v_i\}}{\#\{\text{closed paths of length } n\}} = \frac{e_j^T M^n e_i}{\text{Tr} M^n}
\]
where \( e_i \) is the \( i \)th basis vector.

Now, recall that we have fixed a right eigenvector \( v \) for \( M \) of eigenvalue \( \lambda \), and a left eigenvector \( u \) of the same eigenvalue, normalized so that \( u^T v = 1 \). Since \( M \) is irreducible and aperiodic, by the Perron-Frobenius Theorem we have
\[
\lim_{n \rightarrow \infty} \frac{M^n}{\lambda^n} = vu^T
\]
and in particular
\[
\lim_{n \rightarrow \infty} \frac{e_i^T M^n e_j}{\lambda^n} = e_i^T vu^T e_j = u_i u_j.
\]

As before the measure \( (\pi_i) \) where \( \pi_i = u_i \lambda_i \) is stationary for the Markov chain defined as \( q_{ij} = \frac{\pi_i}{\lambda_i} \), so we consider this Markov chain with the stationary measure as starting distribution. Let \( \nu_{n,m} = \nu_{n-m} \) be the pushforward of the Markov measure on the set of paths of length \( n - m \). Then
\[
\nu_{n,m}(\gamma) = \frac{\pi_i(v_j)}{v_i \lambda^{n-m}} = \frac{u_i v_j}{\lambda^{n-m}}.
\]
Hence
\[ \frac{d\lambda_{n,m}}{d\nu_{n,m}}(\gamma) = \frac{e^T M^m e_i \lambda^{n-m}}{\text{Tr} M^n u_i v_j} \frac{\lambda^n}{\lambda^{m u_i v_j} \text{Tr} M^n} \to 1, \]
as \min\{m, n\} \to \infty. \hfill \Box

We conclude this section by promoting the CLT for displacement from the Markov chain (Theorem 8) to the counting measure \( \lambda_n \).

**Theorem 13.** For any \( a, b \in \mathbb{R} \) with \( a < b \) one has
\[ \lambda_n \left( x : \frac{d(\text{ev}(x)z, z) - nL}{\sigma \sqrt{n}} \in [a, b] \right) \to N_{a,b} \]
as \( n \to \infty \).

**Proof.** Since \( \log n/\sqrt{n} \to 0 \), Theorem 8 (together with Lemma 3) implies
\[ \nu_{n-\log n} \left( x : \frac{d(\text{ev}(x)z, z) - nL}{\sigma \sqrt{n}} \in [a, b] \right) \to N_{a,b}. \]
Now, since \( \nu_{n-\log n} = \nu_{n,\log n} \) and by Proposition 12 and Lemma 4 we get
\[ \|\nu_{n,\log n} - \lambda_{n,\log n}\|_{TV} \to 0 \]
hence
\[ \lambda_{n,\log n} \left( x : \frac{d(\text{ev}(x)z, z) - nL}{\sigma \sqrt{n}} \in [a, b] \right) \to N_{a,b}. \]

Moreover, by the definition of \( \widehat{\lambda} \),
\[ \lambda_n \left( x : \frac{d(\text{ev}(\widehat{x})z, z) - nL}{\sigma \sqrt{n}} \in [a, b] \right) = \lambda_{n,\log n} \left( x : \frac{d(\text{ev}(x)z, z) - nL}{\sigma \sqrt{n}} \in [a, b] \right). \]
Finally, note the since the orbit map \( G \to \mathbb{H}^2 \) is Lipschitz, we have
\[ |d(\text{ev}(x)z, z) - d(\text{ev}(\widehat{x})z, z)| \leq C \log n \]
where \( C \) is the Lipschitz constant. Then using Lemma 3
\[ \lim_{n \to \infty} \lambda_n \left( x : \frac{d(\text{ev}(x)z, z) - nL}{\sigma \sqrt{n}} \in [a, b] \right) = \lim_{n \to \infty} \lambda_n \left( x : \frac{d(\text{ev}(\widehat{x})z, z) - nL}{\sigma \sqrt{n}} \in [a, b] \right) = N_{a,b} \]
which completes the proof. \hfill \Box

### 5. The Gromov Product

The remaining step of our proof is to turn the statement about displacement (Theorem 13) into a statement about translation length. This is done by controlling the Gromov product.

We begin with the following easy computation. Recall that \( \Omega^n \) is the set of all paths of length \( n \) and \( C_n \subset \Omega^n \) is the subset of closed paths.

**Lemma 14.** There is a constant \( D \geq 0 \) such that
\[ 1 \leq \frac{\#(\Omega^n)}{\#(C_n)} \leq D \]
Proof. We know that \( \#(C_n) = \text{Tr } M^n = \lambda^n + \sum_i \lambda_i^n \), where \( \lambda \) is the Perron–Frobenius eigenvalue of \( M \) and \( |\lambda_i| < \lambda \). Also, \( \#(\Omega^n) = \|M^n\|_1 \leq D\lambda^n \) for some \( D \geq 1 \). \( \square \)

Next, we will see that the Gromov product of a random element and its inverse grows slowly in word length. This is our key estimate for relating displacement to translation length.

**Proposition 15.** For any \( \epsilon > 0 \),
\[
\lambda_n(x) : (\text{ev}(x)z, \text{ev}(x)^{-1}z) \geq \epsilon \sqrt{n} \rightarrow 1,
\]
as \( n \rightarrow \infty \).

**Proof.** Fix \( \epsilon > 0 \) and a vertex \( v \) of \( \Gamma \).

Recall that \( \Omega_v \) is the set of one-sided, infinite paths in the graph starting at \( v \). Let us denote as \( P_v := \nu(\cdot | \Omega_v) \) the conditional probability of \( \nu \) given \( \Omega_v \). This is a probability measure on \( \Omega_v \subseteq \Omega \), and is defined on a cylinder set \( C(i_0, i_1, \ldots, i_k) \) as
\[
P_v(C(i_0, i_1, \ldots, i_k)) = \delta_{v,i_0}q_{i_0,i_1}\cdots q_{i_{k−1},i_k}
\]
where \( \delta_{v,i_0} = 1 \) if \( v = i_0 \), and \( \delta_{v,i_0} = 0 \) otherwise and the \( q_{ij} \) are as in (2). Let \( x_n: \Omega_v \rightarrow \Omega_v^n \) be the prefix map which sends an infinite paths starting at \( v \) to its prefix of length \( n \). Define
\[
A = \{ x \in \Omega^* : (\text{ev}(x)z, \text{ev}(x)^{-1}z) \geq \epsilon \sqrt{|x|} \},
\]
where \( |x| \) denotes the length of \( x \). By [10, Section 6] for any vertex \( v \) of \( \Gamma \) one has
\[
P_v((\text{ev}(x_n)z, \text{ev}(x_n)^{-1}z) \geq \epsilon \sqrt{n}) \rightarrow 0.
\]
(Lemma 6.26 of [10] explicitly gives this statement where \( v \) is the “initial vertex” of the directed graph, however the same argument gives the more general result for all vertices.)

Next, consider any path \( p \in \Omega_v^n \) of length \( n \) starting at \( v \). If \( v \) is the \( l \)th vertex of \( \Gamma \) and the terminal endpoint of \( p \) is the \( k \)th vertex, then the \( n \)th step measure of \( p \) is \( P^n_v(p) = \frac{1}{\lambda^n} \frac{v_k}{\nu_k} \). (Here, \( P^n_v \) is the measure supported on \( \Omega_v^n \) which is the pushforward of \( P_v \) under the prefix map \( \Omega_v \rightarrow \Omega_v^n \)).

Then, for any subset \( B \subset \bigcup_{n \geq 1} \Omega_v^n \),
\[
P^n_v(B) = \frac{1}{\lambda^n} \sum_{k=1}^r v_k \#(B \cap \Omega_v^n)k
\]
and
\[
\#B \cap \Omega_v^n = \sum_{k=1}^r \#(B \cap \Omega_v^n)k
\]
where \( \Omega_v^n,w \) is the set of paths of length \( n \) starting at \( v \) and ending at \( w \). Moreover, by [10, Lemma 2.3 (3)], there is a constant \( K > 0 \) such that \( K^{-1} \lambda^n \leq \#\Omega_v^n \leq K\lambda^n \) for any \( n \geq 0 \). Hence, there is a constant \( c > 1 \), depending only on the adjacency matrix \( M \), such that
\[
\frac{1}{c} \frac{\#(B \cap \Omega_v^n)}{\#\Omega_v^n} \leq P^n_v(B) \leq c \cdot \frac{\#(B \cap \Omega_v^n)}{\#\Omega_v^n}.
\]
In particular, we conclude that
\[ \frac{\#(A \cap \Omega^n)}{\#\Omega^n} \leq c \cdot \mathbb{P}_v \left( (ev(x_n)z, ev(x_n)^{-1}z) \geq \epsilon \sqrt{n} \right) \to 0, \]
as \( n \to \infty \).

Hence, by summing over all vertices \( v \in V \)
\[ \frac{\#(A \cap \Omega^n)}{\#\Omega^n} \to 0. \]
Then by combining this with Lemma 14,
\[ \lambda_n(A) = \frac{\#(A \cap C_n)}{\#C_n} \leq \frac{\#(A \cap \Omega^n)}{\#\Omega^n} \cdot \frac{\#\Omega^n}{\#C_n} \to 0. \]

\( \square \)

**Remark 16.** The main estimate (6) in the proof of Proposition 15 can also be obtained via a trick using more recent work of the authors [11]. Using terminology there, one may define a geodesic graph structure \((G, \Gamma)\) by declaring that \( v \) be the initial vertex of \( \Gamma \). Such a structure may not be surjective, but this is not necessary. Then [11, Proposition 5.8] precisely gives the required decay result.

**Proof of Theorem 1.** We can now complete the proof of Theorem 1.

**Proof of Theorem 1.** Let \( d_n := \frac{d(z, ev(x)z) - nL}{\sigma \sqrt{n}} \), \( t_n := \frac{\tau(ev(x)) - nL}{\sigma \sqrt{n}} \), and \( p_n := t_n - d_n = \frac{2(ev(x)z, ev(x)^{-1}z) + O(\delta)}{\sigma \sqrt{n}} \). By the CLT for displacement (Theorem 8), for any \( a < b \)
\[ \lambda_n(x : d_n \in [a, b]) \to N_{a,b}. \]
Moreover, by decay of Gromov products for any \( \epsilon > 0 \) (Proposition 15) we have
\[ \lambda_n(x : |p_n| \geq \epsilon) \to 0 \]
hence by Lemma 3
\[ \lambda_n \left( x : \frac{\tau(ev(x)) - nL}{\sigma \sqrt{n}} \in [a, b] \right) \to N_{a,b}. \]
Finally, by Lemma 6 this implies
\[ \mu_n \left( \gamma : \frac{\tau(\gamma) - nL}{\sigma \sqrt{n}} \in [a, b] \right) \to N_{a,b} \]
which completes the proof. \( \square \)

**6. Generalizations**

While many generalization of Theorem 1 are possible, we record the most immediate one here. The proof is the same as the one given above, using that Lemma 9 holds for convex cocompact actions on \( \text{CAT}(-1) \) spaces [18].

**Theorem 17.** Let \( G \curvearrowright X \) be any convex cocompact action of a closed orientable surface group on a \( \text{CAT}(-1) \) space. Then the conclusion of Theorem 1 holds. The same is true for any convex cocompact free group action on a \( \text{CAT}(-1) \) space \( X \) as long as the lengths of closed geodesics on \( X/G \) are not all contained in \( c\mathbb{Z} \) for some \( c > 0 \).
The condition on lengths of closed geodesics is known to hold whenever \( G \curvearrowright X \) is a discrete action on a \( \text{CAT}(-1) \) space satisfying at least one of the following:

- the limit set \( \Lambda(G) \subseteq \partial X \) has an infinite connected component (in particular when \( G \) is a surface group) \([1]\);
- \( X \) is itself a surface (so that \( X/G \) is a locally \( \text{CAT}(-1) \) surface) \([9]\);
- \( X \) is a rank 1 symmetric space \([13]\).

References


**Department of Mathematics, University of Toronto, 40 St George St, Toronto, ON, Canada,**

*E-mail address*: ilyagekh@gmail.com

**Department of Mathematics, Temple University, 1805 North Broad Street Philadelphia, PA 19122, U.S.A.,**

*E-mail address*: samuel.taylor@temple.edu

**Department of Mathematics, University of Toronto, 40 St George St, Toronto, ON, Canada,**

*E-mail address*: tiozzo@math.toronto.edu