The good pants homology and the Ehrenpreis Conjecture

By Jeremy Kahn and Vladimir Markovic

Abstract

We develop the notion of the good pants homology and show that it agrees with the standard homology on closed surfaces. (Good pants are pairs of pants whose cuffs have the length nearly equal to some large number $R > 0$.) Combined with our previous work on the Surface Subgroup Theorem, this yields a proof of the Ehrenpreis Conjecture.

1. Introduction

Let $S$ and $T$ denote two closed Riemann surfaces. (All closed surfaces in this paper are assumed to have genus at least 2.) The well-known Ehrenpreis Conjecture asserts that for any $K > 1$, one can find finite degree covers $S_1$ and $T_1$, of $S$ and $T$ respectively, such that there exists a $K$-quasiconformal map $f : S_1 \to T_1$. The purpose of this paper is to prove this conjecture. Below we outline the strategy of the proof.

Let $R > 1$, and let $\Pi(R)$ be the hyperbolic pair of pants (with geodesic boundary) whose three cuffs have the length $R$. We define the surface $S(R)$ to be the genus two surface that is obtained by gluing two copies of $\Pi(R)$ along the cuffs with the twist parameter equal to +1. (These are the Fenchel-Nielsen coordinates for $S(R)$.) By $\text{Orb}(R)$ we denote the quotient orbifold of the surface $S(R)$ (the quotient of $S(R)$ by the group of automorphisms of $S(R)$). For a fixed $R > 1$, we sometimes refer to $\text{Orb}(R)$ as the model orbifold. The following theorem is the main result of this paper.

**Theorem 1.1.** Let $S$ be a closed hyperbolic Riemann surface. Then for every $K > 1$, there exists $R_0 = R_0(K, S) > 0$ such that for every $R > R_0$, there are finite covers $S_1$ and $O_1$ of the surface $S$ and the model orbifold $\text{Orb}(R)$ respectively, and a $K$-quasiconformal map $f : S_1 \to O_1$.

The Ehrenpreis Conjecture is an immediate corollary of this theorem.

---

JK is supported by NSF grant number DMS-0905812.
© 2015 Department of Mathematics, Princeton University.
Corollary 1.1. Let $S$ and $T$ denote two closed Riemann surfaces. For any $K > 1$, one can find finite degree covers $S_1$ and $T_1$ of $S$ and $T$ respectively, such that there exists a $K$-quasiconformal map $f : S_1 \to T_1$.

Proof. Fix $K > 1$. It follows from Theorem 1.1 that for $R$ large enough, there exist

(i) finite covers $S_1$, of $S$ and $T$ respectively;
(ii) finite covers $O_1$ and $O'_1$ of Orb($R$);
(iii) $\sqrt{K}$-quasiconformal mappings $f : S_1 \to O_1$ and $g : T_1 \to O'_1$.

Let $O_2$ denote a common finite cover of $O_1$ and $O'_1$. (Such $O_2$ exists since $O_1$ and $O'_1$ are covers of the same orbifold Orb($R$).) Then there are finite covers $S_2$ and $T_2$, of $S_1$ and $T_1$, respectively, and the $\sqrt{K}$-quasiconformal maps $\tilde{f} : S_2 \to O_2$ and $\tilde{g} : T_2 \to O_2$, that are the lifts of $f$ and $g$. Then $\tilde{g}^{-1} \circ \tilde{f} : S_2 \to T_2$ is $K$-quasiconformal map, which proves the corollary. \qed

In the remainder of the paper we prove Theorem 1.1. This paper builds on our previous paper [4], where we used immersed skew pants in a given hyperbolic 3-manifold to prove the Surface Subgroup Theorem. We note that Lewis Bowen [1] was the first to attempt to build finite covers of Riemann surfaces by putting together immersed pairs of pants. We also note that it follows from the work of Danny Calegari [2] that the pants homology is equal to the standard homology. This means that every sum of closed curves on a closed surface $S$ that is zero in the standard homology $H_1(S)$ is the boundary of a sum of immersed pairs of pants in $S$.

We are grateful to Lewis Bowen for carefully reading the manuscript and suggesting numerous improvements and corrections. The second named author would like to acknowledge that the first named author has done most of the work in the second part of the paper concerning the Correction Theory.

Outline. In our previous paper [4] we proved a theorem very similar to Theorem 1.1, namely that given a closed hyperbolic 3-manifold $M^3$, and $K > 1$, $R > R_0(K, M^3)$, we can find a finite cover $O_1$ of Orb($R$) and a map $f : O_1 \to M^3$ that lifts to a map $\tilde{f} : \mathbb{H}^2 \to \mathbb{H}^3$ such that $\partial \tilde{f} : \partial \mathbb{H}^2 \to \partial \mathbb{H}^3$ has a $K$-quasiconformal extension. We proved that theorem by finding a large collection of “skew pairs of pants” whose cuffs have complex half-lengths close to $R$, and which are “evenly distributed” around every closed geodesic that appears as a boundary.

We then assemble these pants by (taking two copies of each and then) pairing off the pants that have a given geodesic as boundary, so that the resulting complex twist-bends (or reduced Fenchel-Nielsen coordinates) are close to 1. We can then construct a cover $O_1$ of Orb($R$) and a function $f : O_1 \to M$.
whose image is the closed surface that results from this pairing. The function $f$ will then have the desired property.

We would like to proceed in the same manner in dimension 2, that is, when a 3-manifold $M^3$ is replaced with a closed surface $S$. We can, as before, find a collection of good immersed pants (with cuff length close to $R$) that is “evenly distributed” around each good geodesic (of length close to $R$) that appears as boundary. If and when we can assemble the pants to form a closed surface with (real) twists close to 1, we will have produced a $K$-quasiconformal immersion of a cover of Orb($R$) into $S$.

There is only one minor detail: the unit normal bundle of a closed geodesic in $S$ has two components. In other words, an immersed pair of pants that has a closed geodesic $\gamma$ as a boundary can lie on one of the two sides of $\gamma$. If, in our formal sum of pants, we find we have more pants on one side of $\gamma$ than the other, then we have no chance to form a closed surface out of this formal sum of pants. It is for this very reason that the Ehrenpreis Conjecture is more difficult to prove than the Surface Subgroup Theorem.

Because our initial collection is evenly distributed, there are almost the same number of good pants on both sides of any good geodesic, so it is natural to try to adjust the number of pairs of pants so that it is balanced (with the same number of pants on both sides of each geodesic). This leads us to look for a “correction function” $\phi$ from formal sums of (normally) oriented closed geodesics (representing the imbalance) to formal sums of good pants, such that the boundary of $\phi(X)$ is $X$.

The existence of this correction function then implies that “good boundary is boundary good”; that is, any sum of good geodesics that is a boundary in $H_1(S)$ (the first homology group of $S$) is the boundary of a sum of good pants. Thus we define the good pants homology to be formal sums of good geodesics (with length close to $R$) modulo boundaries of formal sums of good pants. We would like to prove that the good pants homology is the standard homology on good curves.

The natural approach is to show that any good curve is homologous in good pants homology to a formal sum of $2g$ particular good curves that represent a basis in $H_1(S)$. ($g$ is the genus of $S$.) That is, we want to show that there are $\{h_1, \ldots, h_{2g}\}$ good curves that generate $H_1(S)$ (here $H_1(S)$ is taken with rational coefficients) such that every good curve $\gamma$ is homologous in the good pants homology to a formal sum $\sum a_i h_i$ for some $a_i \in \mathbb{Q}$. Then any sum of good curves is likewise homologous to a sum of good generators $h_i$, but if the original sum of good curves is zero in $H_1(S)$, then the corresponding sum of $h_i$’s is zero as well.

To prove the Good Correction Theorem, we must first develop the theory of good pants homology. Let $*$ denote a base point on $S$. For $A \in \pi_1(S, *) \setminus \{\text{id}\}$,
we let \([A]\) denote the closed geodesic freely homotopic to \(A\). Our theory begins
with the Algebraic Square Lemma, which states that, under certain geometric
conditions,

\[
\sum_{i,j=0,1} (-1)^{i+j} [A_i U B_j V] = 0
\]

in the good pants homology. (The curves \([A_i U B_j V]\) must be good curves, the
words \(A_i U B_j V\) reasonably efficient, and the length of \(U\) and \(V\) sufficiently
large.) This then permits us to define, for \(A,T \in \pi_1(S,\ast)\),

\[
A_T = \frac{1}{2} ([TAT^{-1}U] - [TA^{-1}T^{-1}U]),
\]

where \(U\) is fairly arbitrary. Then \(A_T\) in good pants homology is independent
of the choice of \(U\).

We then show through a series of lemmas that \((XY)_T = X_T + Y_T\) in good
pants homology, and therefore

\[
X_T = \sum \sigma(j)(g_{i_j})_T,
\]

where by \(X = g_{i_1}^{\sigma(1)} \cdots g_{i_k}^{\sigma(k)}\) we have written \(X\) as a product of generators.
(Here \(g_1, \ldots, g_{2g}\) are the generators for \(\pi_1(S,\ast)\) and \(\sigma(j) = \pm 1\).) With a little
more work we can show

\[
[X] = \sum \sigma(j)(g_{i_j})_T
\]
as well, and thus we can correct any good curve to good generators.

We are then finished except for one last step: We must show that our
correction function, which gives an explicit sum of pants with a given boundary,
is well behaved in that it maps sums of curves, with bounded weight on each
curve, to sums of pants, with bounded weight on each pair of pants. We
call such a function semirandom, because if we pick a curve at random, the
expected consumption of a given pair of pants is not much more than if we
picked the pair of pants at random.

We define the correction function implicitly, through a series of lemmas,
each of which asserts the existence of a formal sum of pants with given bound-
dary, and which is in principle constructive. The notion of being semirandom is
sufficiently natural to permit us to say that the basic operations, such as
the group law, or forming \([A]\) from \(A\), as well as composition and formal addition,
are all semirandom. So in order to verify that our correction function is semi-
random, we need only to go through each lemma observing that the function
we have defined is built out of the functions that we have previously defined
using the standard operations that we have proved are semirandom.

To make the paper as easy to read as possible, we have relegated the
verification of semi-randomness to the “Randomization remarks” that follow
our homological lemmas and that use the notation and results (that basic con-
structions are semirandom) that we have placed in the appendix. We strongly
recommend that the reader skip over these Randomization remarks in the first reading and to interpret the word “random” in the text to simply mean “arbitrary.”

A word on notation. When we use the letter K to denote a constant then we mean a universal constant or if K depends on parameters X, Y, Z, . . . then we write K(X, Y, Z, . . .) or we may leave out some of the parameters. In sections where we fixed certain parameters we may leave out the dependence of constants on these parameters.

2. Constructing good covers of a Riemann surface

2.1. The reduced Fenchel-Nielsen coordinates and the model orbifolds. Let S0 be an oriented closed topological surface with a given pants decomposition C, where C is a maximal collection of disjoint simple closed curves that cut S0 into the corresponding pairs of pants. We will say that C makes S0 into a panted surface.

Denote by C* the set of oriented curves from C. (Each curve in C is taken with both orientations.) The set of pairs (Π, C*), where Π is a pair of pants from the pants decomposition and C* ∈ C* is an oriented boundary cuff of Π, is called the set of marked pants and is denoted by Π(S0). For C ∈ C, there are exactly two pairs (Πi, C∗i) ∈ Π(S0), i = 1, 2, such that C∗1 and C∗2 are the two orientations on C. (Note that Π1 and Π2 may agree as pairs of pants.)

Let (S, C) be a panted Riemann surface. Then for every cuff C ∈ C, we can define the reduced Fenchel-Nielsen coordinates (hl(C), s(C)) from [4]. Here hl(C) is the half-length of the geodesic homotopic to C, and s(C) ∈ R/hl(C)Z is the reduced twist parameter that lives in the circle R/|hl(C)|Z. (When we write s(C) = x ∈ R, we really mean s(C) ≡ x mod(hl(C))Z.) The following theorem was proved in [4] (see Theorem 2.1 and Corollary 2.1 in [4]).

Theorem 2.1. There exist constants 0 < R > 0 such that the following holds. Let S denote a panted Riemann surface whose reduced Fenchel-Nielsen coordinates satisfy the inequalities

$$|hl(C) - R| < \varepsilon \quad \text{and} \quad |s(C) - 1| < \frac{\varepsilon}{R},$$

for some $\varepsilon > \varepsilon > 0$ and $R > \hat{R}$. Then there exist a marked surface MR, with the reduced Fenchel-Nielsen coordinates hl(C) = R and s(C) = 1, and a K-quasiconformal map $f : S \to M_R$, where

$$K = \frac{\hat{\varepsilon} + \varepsilon}{\varepsilon - \varepsilon}.$$
surface $M_R$ with reduced Fenchel-Nielsen coordinates $hl(C) = R$ and $s(C) = 1$ is a finite cover of $\text{Orb}(R)$.

2.2. A proof of Theorem 1.1. Below we state the theorem that is then used to prove Theorem 1.1.

**Theorem 2.2.** Let $S$ denote a closed Riemann surface, and let $\varepsilon > 1$. There exists $R(S, \varepsilon) > 1$ such that for every $R > R(S, \varepsilon)$, we can find a finite cover $S_1$ of $S$ that has a pants decomposition whose reduced Fenchel-Nielsen coordinates satisfy the inequalities

$$|hl(C) - R| < \varepsilon \quad \text{and} \quad |s(C) - 1| < \frac{\varepsilon}{R}.$$

This theorem will be proved at the end of section.

2.3. The set of immersed pants in a given Riemann surface. From now on $S = \mathbb{H}^2/\mathcal{G}$ is a fixed closed Riemann surface and $\mathcal{G}$ a suitable Fuchsian group. By $\Gamma$ we denote the collection of oriented closed geodesics in $S$. By $-\gamma$ we denote the opposite orientation of an oriented geodesic $\gamma \in \Gamma$. We sometimes write $\gamma^* \in \Gamma$ to emphasize a choice of orientation.

Let $\Pi_0$ denote a hyperbolic pair of pants. (That is, $\Pi_0$ is equipped with a hyperbolic metric such that the cuffs of $\Pi_0$ are geodesics.) Let $f : \Pi_0 \to S$ be a local isometry. (Such an $f$ must be an immersion.) We say that $\Pi = (f, \Pi_0)$ is an immersed pair of pants in $S$. The set of all immersed pants in $S$ is denoted by $\Pi$. The set of such pairs $(\Pi, \gamma)$ is called the set of marked immersed pants and denoted by $\Pi^*$. The half-length $hl(\gamma)$ associated to the cuff $\gamma$ of $\Pi$ is defined as the half-length $hl(C)$ associated to the cuff $C$ of $\Pi_0$.

Let $\gamma \in \Gamma$ be a closed oriented geodesic in $S$. Denote by $N^1(\gamma)$ the unit normal bundle of $\gamma$. Elements of $N^1(\gamma)$ are pairs $(p, v)$, where $p \in \gamma$ and $v$ is a unit vector at $p$ that is orthogonal to $\gamma$. The bundle $N^1(\gamma)$ is a differentiable manifold that has two components, which we denote by $N^1_+(\gamma)$ and $N^1_-(\gamma)$ (the right-hand side and the left-hand side components). Each component inherits the metric from the geodesic $\gamma$, and both $N^1_+(\gamma)$ and $N^1_-(\gamma)$
are isometric (as metric spaces) to the circle of length $2\text{hl}(\gamma)$. By $\text{dis}$ we denote the corresponding distance functions on $N^1_+(\gamma)$ and $N^1_-(\gamma)$.

Let $(p, v) \in N^1(\gamma)$, and denote by $(p_1, v_1) \in N^1(\gamma)$ the pair such that $(p, v)$ and $(p_1, v_1)$ belong to the same component of $N^1(\gamma)$, and $\text{dis}(p, p_1) = \text{hl}(\gamma)$. Set $A(p, v) = (p_1, v_1)$. Then $A$ is an involution that leaves invariant each component of $N^1(\gamma)$. Define the bundle $N^1(\sqrt{\gamma}) = N^1(\gamma)/A$. The two components are denoted by $N^1_+(\sqrt{\gamma})$ and $N^1_-(\sqrt{\gamma})$, and both are isometric (as metric spaces) to the circle of length $\text{hl}(\gamma)$. The disjoint union of all such bundles is denoted by $N^1(\sqrt{\Gamma})$.

We now define the foot of a pair of pants. Let $\Pi \in \Pi$ be an immersed pants and $f : \Pi_0 \to \Pi$ the corresponding local isometry. Let $C^*$ denote an oriented cuff of $\Pi_0$ and $\gamma = f(C^*)$. Let $C_1$ and $C_2$ denote the other two cuffs of $\Pi_0$, and let $p_1', p_2' \in C^*$ denote the two points that are the feet of the shortest geodesic segments in $\Pi_0$ that connect $C$ and $C_1$, and $C$ and $C_2$, respectively. Let $v_1'$ denote the unit vector at $p_1'$ that is orthogonal to $C$ and points towards the interior of $\Pi_0$. We define $v_2'$ similarly. Set $(p_1, v_1) = f_*(p_1', v_1')$ and $(p_2, v_2) = f_*(p_2', v_2')$. Then $(p_1, v_1)$ and $(p_2, v_2)$ are in the same component of $N^1(\gamma)$, and the points $p_1$ and $p_2$ separate $\gamma$ into two intervals of length $\text{hl}(\gamma)$. Therefore, the vectors $(p_1, v_1)$ and $(p_2, v_2)$ represent the same point $(p, v) \in N^1(\sqrt{\Gamma})$, and we set

$$\text{foot}(\Pi, \gamma) = (p, v) \in N^1(\sqrt{\gamma}).$$

We call the vector $(p, v)$ the foot of the immersed pair of pants $\Pi$ at the cuff $\gamma$. This defines the map

$$\text{foot} : \Pi^* \to N^1(\sqrt{\Gamma}).$$

2.4. Measures on pants and the $\partial$ operator. By $\mathcal{M}(\Pi)$ we denote the space of real valued Borel measures with finite support on the set of immersed pants $\Pi$, and likewise, by $\mathcal{M}(N^1(\sqrt{\Gamma}))$ we denote the space of real valued Borel measures with compact support on the manifold $N^1(\sqrt{\Gamma})$. (A measure from $\mathcal{M}^+(N^1(\sqrt{\Gamma}))$ has a compact support if and only if its support is contained in at most finitely many bundles $N^1(\sqrt{\gamma}) \subset N^1(\sqrt{\Gamma})$.) By $\mathcal{M}^+(\Pi)$ and $\mathcal{M}^+(N^1(\sqrt{\Gamma}))$, we denote the corresponding spaces of positive measures.

We define the operator

$$\partial : \mathcal{M}(\Pi) \to \mathcal{M}(N^1(\sqrt{\Gamma}))$$

as follows. The set $\Pi$ is a countable set, so every measure from $\mu \in \mathcal{M}(\Pi)$ is determined by its value $\mu(\Pi)$ on every $\Pi \in \Pi$. Let $\Pi \in \Pi$, and let $\gamma_i \in \Gamma$, $i = 0, 1, 2$, denote the corresponding oriented geodesics so that $(\Pi, \gamma_i) \in \Pi^*$. Let $\alpha_{\Pi} \in \mathcal{M}(N^1(\sqrt{\Gamma}))$ be the atomic measure supported at the point $\text{foot}(\Pi, \gamma_i) \in N^1(\sqrt{\gamma_i})$, where the mass of the atom is equal to 1. Let

$$\alpha_\Pi = \sum_{i=0}^2 \alpha_{\Pi}^i,$$
and define
\[ \hat{\partial} \mu = \sum_{\Pi \in \Pi} \mu(\Pi) \alpha^\Pi. \]

We call this the \( \hat{\partial} \) operator on measures. If \( \mu \in \mathcal{M}^+(\Pi) \), then
\[ \hat{\partial} \mu \in \mathcal{M}^+(N^1(\sqrt{T})), \]
and the total measure of \( \hat{\partial} \mu \) is three times the total measure of \( \mu \).

We recall the notion of equivalent measures from Section 3 in [4]. Let \( (X, d) \) be a metric space. By \( \mathcal{M}^+(X) \) we denote the space of positive, Borel measures on \( X \) with compact support. For \( A \subseteq X \) and \( \delta > 0 \), let
\[ N_\delta(A) = \{ x \in X : \text{there exists } a \in A \text{ such that } d(x, a) \leq \delta \}, \]
be the \( \delta \)-neighbourhood of \( A \).

**Definition 2.1.** Let \( \mu, \nu \in \mathcal{M}^+(X) \) be two measures such that \( \mu(X) = \nu(X) \), and let \( \delta > 0 \). Suppose that for every Borel set \( A \subseteq X \) we have \( \mu(A) \leq \nu(N_\delta(A)) \). Then we say that \( \mu \) and \( \nu \) are \( \delta \)-equivalent measures.

**Remark.** We observe that this definition is symmetric because \( \nu(A) \leq \mu(N_\delta(A)) \) whenever \( \mu(X \setminus N_\delta(A)) \leq \nu(N_\delta(X \setminus N_\delta(A)) \).

In our applications \( X \) will be either a 1-torus (a circle) or \( \mathbb{R} \). In this case, \( \mu(A) \leq \nu(N_\delta(A)) \) for all Borel sets \( A \) if it holds for all intervals \( A \). We recall that if \( \mu \) and \( \nu \) are discrete and integer valued measures that are \( \varepsilon \)-equivalent, then there is a “matching” between \( \mu \) and \( \nu \) that matches each point \( x \) to a point within \( \varepsilon \) of \( x \). In other words, letting \( E_n \) be \( \{1, 2, \ldots, n\} \) with the counting measure, if \( \mu = f_*E_n \) and \( \nu = g_*E_n \) and \( \mu \) and \( \nu \) are \( \varepsilon \)-equivalent, then we can find \( \sigma : E_n \rightarrow E_n \) such that \( d(f(k), g(\sigma(k))) \leq \varepsilon \) for each \( k \in E_n \). This holds when \( \mu \) and \( \nu \) are measures on any metric space (by the Hall’s Marriage Theorem) and is even more elementary when the metric space is a 1-torus (i.e., a circle) or an interval.

We observe that if \( \mu \) and \( \nu \) are \( \varepsilon_1 \)-equivalent and \( \nu \) and \( \rho \) are \( \varepsilon_2 \)-equivalent, then \( \mu \) and \( \rho \) are \( (\varepsilon_1 + \varepsilon_2) \)-equivalent.

Let \( \gamma \in \Gamma \) and \( \alpha \in \mathcal{M}(N^1(\sqrt{T})) \). The bundle \( N^1(\sqrt{T}) \) has the two components \( N^1_+(\sqrt{T}) \) and \( N^1_-(-\sqrt{T}) \) (the right-hand and left-hand side components), each isometric to the circle of length \( \text{hl}(\gamma) \). The restrictions of \( \alpha \) to \( N^1_+(-\sqrt{T}) \) and \( N^1_-(-\sqrt{T}) \) are denoted by \( \alpha_+ \) and \( \alpha_- \) respectively. In particular, by \( \hat{\partial}_+ \mu \) and \( \hat{\partial}_- \mu \) we denote the decomposition of the measure \( \hat{\partial} \mu \).

**Definition 2.2.** Fix \( \gamma \in \Gamma \), and let \( \alpha, \beta \in \mathcal{M}^+(N^1(\sqrt{T})) \). We say that \( \alpha \) and \( \beta \) are \( \delta \)-equivalent if the pairs of measures \( \alpha_+ \) and \( \beta_+ \), and \( \alpha_- \) and \( \beta_- \), are respectively \( \delta \)-equivalent. Also, by \( \lambda(\gamma) \in \mathcal{M}^+(N^1(\sqrt{T})) \), we denote the measure whose components \( \lambda_+(\gamma) \) and \( \lambda_-(\gamma) \) are the standard Lebesgue measures.
on the metric spaces $N^1(\sqrt{\gamma})^+$ and $N^1(\sqrt{\gamma})^-$, respectively. In particular, the measure $\lambda(\gamma)$ is invariant under the full group of isometries of $N^1(\sqrt{\gamma})$.

Let $\varepsilon, R > 0$. By $\Gamma_{\varepsilon,R} \subset \Gamma$ we denote the closed geodesics in the Riemann surface $S$ whose half-length is in the interval $[R-\varepsilon, R+\varepsilon]$. We define $\Pi_{\varepsilon,R} \subset \Pi$, as the set of immersed pants whose cuffs are in $\Gamma_{\varepsilon,R}$. We will often call $\Gamma_{\varepsilon,R}$ the set of “good curves” and $\Pi_{\varepsilon,R}$ the set of “good pants.” Our aim is to prove the following theorem, which in turn yields the proof of Theorem 2.2 stated above.

We adopt the following convention. In the rest of the paper by $P(R)$ we denote a polynomial in $R$ whose degree and coefficients depend only on the choice of $\varepsilon$ and the surface $S$.

**Theorem 2.3.** Let $\varepsilon > 0$. There exists $q > 0$ (depending only on the surface $S$ and $\varepsilon$) so that for every $R > 0$ large enough, there exists a measure $\mu \in M^+(\Pi_{\varepsilon,R})$ with the following properties. Let $\gamma \in \Gamma_{\varepsilon,R}$, and let $\hat{\partial}\mu(\gamma)$ denote the restriction of $\hat{\partial}\mu$ to $N^1(\sqrt{\gamma})$. If $\hat{\partial}\mu(\gamma)$ is not the zero measure, then there exists a constant $K_\gamma > 0$ such that the measures $\hat{\partial}\mu(\gamma)$ and $K_\gamma \lambda(\gamma)$ are $P(R)e^{-qR}$-equivalent.

**Remark.** We say that $\alpha \in M(N^1(\sqrt{\gamma}))$ is $\delta$ symmetric (for $\delta > 0$) if for every isometry $\iota : N^1(\sqrt{\gamma}) \rightarrow N^1(\sqrt{\gamma})$, the measures $\alpha$ and $\iota_*\alpha$ are $\delta$-equivalent. If $\alpha$ and $K_\gamma \lambda(\gamma)$ are $\delta$-equivalent, then $\alpha$ is $2\delta$ symmetric because $\lambda(\gamma)$ is 0 symmetric.

The proof of Theorem 2.2 follows from Theorem 2.3 in the same way as it was done in Section 3 in [4]. The brief outline is as follows. We may assume that the measure $\mu$ from the above theorem has integer coefficients. Then we may think of $\mu$ as a formal sum of immersed pants such that the restriction of the measure $\hat{\partial}\mu$ on any $N^1(\sqrt{\gamma})$ is $P(R)e^{-qR}$-equivalent with some multiple of the Lebesgue measure (unless the restriction is the zero measure). Considering $\mu$ as the multiset (formally one may use the notion of a labelled collection of immersed pants) we can then define a pairing between marked immersed pants, such that two marked pants $(\Pi_1, \gamma_1)$ and $(\Pi_2, \gamma_2)$ are paired if $\gamma_1 = -\gamma_2$, and the corresponding twist parameter between these two pairs is $P(R)e^{-pR}$ close to $+1$ for some universal constant $p > 0$. After gluing all the marked pants we have paired, we obtain the finite cover from Theorem 2.2.

### 3. Equidistribution and self-correction

In this section we introduce the Equidistribution Theorem (Theorem 3.1) and the Correction Theorem (Theorem 3.4) and we use them to prove Theorem 2.3. Theorem 3.1 follows from our previous work [4] and provides us with an evenly distributed collection of good pants. The Correction Theorem allows us to correct the slight imbalance (as described in the introduction) that may be found in the original collection of pants.
3.1. Generating essential immersed pants in $S$. Let us first describe how we generate a pair of pants from a $\Theta$-graph. Let $\Pi$ denote a pair of pants whose three cuffs have the same length, and let $\omega : \Pi \to \Pi$ denote the standard (orientation preserving) isometry of order three that permutes the cuffs of $\Pi$. Let $a$ and $b$ be the fixed points of $\omega$. Let $\gamma_0 \subset \Pi$ denote a simple oriented geodesic arc that starts at $a$ and terminates at $b$. Set $\omega^i(\gamma_0) = \gamma_i$. The union of two different arcs $\gamma_i$ and $\gamma_j$ is a closed curve in $\Pi$ homotopic to a cuff. One can think of the union of these three segments as the spine of $\Pi$. Moreover, there is an obvious projection from $\Pi$ to the spine $\gamma = \gamma_0 \cup \gamma_1 \cup \gamma_2$, and this projection is a homotopy equivalence.

Let $p$ and $q$ be two (not necessarily) distinct points in $S$, and let $\alpha_0$, $\alpha_1$, and $\alpha_2$ denote three distinct oriented geodesic arcs, each starting at $p$ and terminating at $q$. We let $\alpha = \alpha_0 \cup \alpha_1 \cup \alpha_2$. (We call $\alpha$ a $\Theta$-graph.) Let $i(\alpha_j) \in T_p^1 S$ and $t(\alpha_j) \in T_q^1 S$ denote the initial and terminal unit tangent vectors to $\alpha_j$ at $p$ and $q$ respectively. Suppose that the triples of vectors $(i(\alpha_0), i(\alpha_1), i(\alpha_2))$ and $(t(\alpha_0), t(\alpha_1), t(\alpha_2))$ have opposite cyclic orders on the unit circle.

We define the map $f : \Pi \to S$ by first projecting the pants $\Pi$ onto its spine $\gamma$ and then by mapping $\gamma_j$ onto $\alpha_j$ by a map that is a local (orientation preserving) homeomorphism. Then the induced conjugacy class of maps $f_* : \pi_1(\Pi) \to \pi_1(S)$ is injective.

Moreover, we can homotop the map $f$ to an immersion $g : \Pi \to S$ as follows. We can write the pants $\Pi$ as a (nondisjoint) union of three strips $G_0, G_1, G_2$, where each $G_i$ is a fattening of the geodesic arc $\gamma_i$. Then we define a map $g_i : G_i \to S$ to be a local homeomorphism on each $G_i$ by extending the restriction of the map $f$ on $\gamma_i$. The condition on the cyclic order of the $\alpha_i$’s at the two vertices enables us to define $g_i$ and $g_j$ on $G_i$ and $G_j$ respectively, so that $g_i = g_j$ on $G_i \cap G_j$. Then we set $g = g_i$.

We say that $g : \Pi \to S$ is the essential immersed pair of pants generated by the three geodesic segments $\alpha_0, \alpha_1$, and $\alpha_2$.

Often we will be given two geodesic segments, say $\alpha_0$ and $\alpha_1$, and then find a third geodesic segment $\alpha_2$ so as to obtain an essential immersed pair of pants. We then say that $\alpha_2$ is a third connection. In this paper we will often be given a closed geodesic $C$ on a Riemann surface $S$, with two marked points $p, q \in C$. Then every geodesic arc $\alpha$ between $p$ and $q$ that meets $p$ and $q$ at the same sides of $C$ will be called a third connection, since then $C$ and $\alpha$ generate an immersed pair of pants as described above. In particular, this represents an efficient way of generating pants that contain a given closed geodesic $C$ as its cuff.

3.2. Preliminary lemmas. Let $T^1 \mathbb{H}^2$ denote the unit tangent bundle. Elements of $T^1(\mathbb{H}^2)$ are pairs $(p, u)$, where $p \in \mathbb{H}^2$ and $u \in T^1_p \mathbb{H}^2$. Sometimes we write $u = (p, u)$ and refer to $u$ as a unit vector in $T^1_p \mathbb{H}^2$. By $T^1 S$ we denote
the unit tangent bundle over $\mathbf{S}$. For $u, v \in T^1_u \mathbb{H}^2$, we let $\Theta(u, v)$ denote the unoriented angle between $u$ and $v$. The function $\Theta$ takes values in the interval $[0, \pi]$.

For $L, \varepsilon > 0$, and $(p, u), (q, v) \in T^1 \mathbf{S}$, we let $\text{Conn}_{\varepsilon, L}((p, u), (q, v))$ be the set of unit speed geodesic segments $\gamma : [0, l] \to \mathbf{S}$ such that

- $\gamma(0) = p$ and $\gamma(l) = q$;
- $|l - L| < \varepsilon$;
- $\Theta(u, \gamma'(0)), \Theta(v, \gamma'(l)) < \varepsilon$.

The next lemma will be referred to as the Connection Lemma. It provides a good lower bound on the size of the connection set we have just defined. This lemma also follows from discussion in the appendix.

**Lemma 3.1.** Given $\varepsilon > 0$, we can find $L_0 = L_0(\varepsilon, \mathbf{S})$ such that for any $L > L_0$, and any two vectors $(p, u)$ and $(q, v)$, the set $\text{Conn}_{\varepsilon, L}((p, u), (q, v))$ is nonempty and

$$\left| \text{Conn}_{\varepsilon, L}((p, u), (q, v)) \right| \geq e^{L - L_0}.$$

**Proof.** By $\text{dis}$ we denote a distance function on $T^1 \mathbf{S}$. (We define $\text{dis}$ explicitly later in the paper.) The unit tangent bundle $T^1 \mathbf{S}$ is naturally identified with $G \backslash \text{PSL}(2, \mathbb{R})$, where $G$ is a lattice. Then we can find a neighbourhood $U$ of the identity in $\text{PSL}(2, \mathbb{R})$ so that if $(q, v) \in T^1 \mathbf{S} = G \backslash \text{PSL}(2, \mathbb{R})$ and $x \in U$, then $\text{dis}((q, v), (q, v) \cdot x) < \frac{\varepsilon}{16}$. We let $\psi : U \to [0, \infty)$ be a $C^{\infty}$ function with compact support in $U$, with $\int_U \psi = 1$. For any $(q, v) \in T^1 \mathbf{S}$, we let $\mathcal{N}_U(q, v) = \{(q, v) \cdot s : s \in U\}$, and we let $\psi_{(q, v)}((q, v) \cdot s) = \psi(s)$ on $\mathcal{N}_U(q, v)$. (If $\varepsilon$ is small, then $s \mapsto (q, v) \cdot s$ is injective.) The $C^k$ norm of $\psi_{(q, v)}$ is independent of $(q, v)$.

Let $g_t : T^1 \mathbf{S} \to T^1 \mathbf{S}$ be the geodesic flow. By uniform mixing for uniformly $C^{\infty}$ functions on $\mathbf{S}$,

$$\int_{T^1 \mathbf{S}} \psi_{(q, v)}(g_t(x, w)) \psi_{(p, u)}(x, w) d(x, w) \to 1,$$

uniformly in $(p, u)$ and $(q, v)$, as $t \to \infty$. (We always assume that the Liouville measure is normalized so that the total measure of the tangent bundle is one.) If $\psi_{(q, v)}(g_t(x, w)) \psi_{(p, u)}(x, w) > 0$, then $(x, w) \in \mathcal{N}_U(p, u)$ and $g_t(x, w) \in \mathcal{N}_U(q, v)$.

The segment $g_{[0, t]}(x, w)$ is $\varepsilon$-nearly homotopic (see the definition after this proof) to a unique geodesic segment $\alpha$ connecting $p$ and $q$. The reader can verify that $\alpha \in \text{Conn}_{\varepsilon, L}((p, u), (q, v))$. We let $E_{\alpha, t} \subset \mathcal{N}_U(p, u)$ be the set of $(x, w)$ for which $g_t(x, w) \in \mathcal{N}_U(q, v)$, and $g_{[0, t]}(x, w)$ is $\varepsilon$-homotopic to $\alpha$. Then
by \((1)\),
\[
\sum_{\alpha} \int_{E_{\alpha,t}} \psi_{(q,v)}(g_t(x,w))\psi_{(p,u)}(x,w) \, d(x,w) = 1 + o(1).
\]

On the other hand, we can easily verify that
\[
V(E_{\alpha,t}) \leq K(\psi) e^{-t}.
\]
(Here \(V(E_{\alpha,t})\) is the volume of \(E_{\alpha,t}\).) Hence
\[
\int_{E_{\alpha,t}} \psi_{(q,v)}(g_t(x,w))\psi_{(p,u)}(x,w) \, d(x,w) \leq \int_{E_{\alpha,t}} (\sup_{U} \psi)^2 \, d(x,w)
\]
\[
\leq K(\psi) e^{-t}.
\]
So the number of good \(\alpha\) is at least \(K(\psi)e^t\), as long as \(t\) is large given \(S\) and \(\epsilon\).

\(\Box\)

**Definition 3.1.** Let \(E \geq 0\). We say that two paths \(A\) and \(B\) in \(\mathbb{H}^2\) are \(E\)-nearly homotopic if the distance between the endpoints of \(A\) and \(B\) is at most \(E\). Two paths on the closed surface \(S\) are \(E\)-nearly homotopic if they have lifts to \(\mathbb{H}^2\) that are \(E\)-nearly homotopic.

The following lemma gives the estimate for the number of good pants that are bounded by a given (good) cuff.

**Lemma 3.2.** Let \(0 < \epsilon < 1\). We let \(\gamma \in \Gamma_{\epsilon,R}\) and let \(\Pi_{\epsilon,R}(\gamma)\) denote the set of pants in \(\Pi_{\epsilon,R}\) that contain \(\gamma\) as a cuff. Then
\[
|\Pi_{\epsilon,R}(\gamma)| \asymp R e^R,
\]
where the constant for \(\asymp\) depends only on \(S\) and \(\epsilon\).

**Proof.** For the upper bound, let \(F_{\gamma}\) denote a set of \([2R]\) evenly distributed points on \(\gamma\). If \(\Pi \in \Pi_{\epsilon,R}\) and if \(\gamma \in \partial \Pi\), we let \(\alpha\) be the geodesic segment in \(\Pi\) that is orthogonal to \(\gamma\) at its endpoints and simple on \(\Pi\). Then we let \(\alpha'\) be a geodesic segment connecting two points of \(F_{\gamma}\), such that \(\alpha'\) is \(\frac{1}{2}\)-nearly homotopic to \(\alpha\), and hence \(l(\alpha') \leq l(\alpha) + 1\). It can easily be verified that the length of \(\alpha\) is at most \(R + 9\), thus the length of \(\alpha'\) is at most \(R + 10\).

We leave it to the reader to verify that the number of geodesic segments of length at most \(L\) that connect two given points on the closed surface \(S\) is at most \(e^{R+\text{diam}(S)/\text{Area}(S)}\).

If two pants produce identical \(\alpha'\), then the two pants are the same. Between any two points of \(F_{\gamma}\) there are at most \(N e^R\) such arcs \(\alpha'\) (the constant \(N\) depends only on \(S\)), and the endpoints of \(\alpha'\) are \(R \pm 1\) apart, so the total number of such arcs is at most \(10NRe^R\).

For the lower bound: By Lemma 3.1 we can find \(N(S, \epsilon)e^R\) geodesic segments \(\tilde{\alpha}\) connecting two given diametrically opposite points of \(\gamma\), of length within \(\frac{\epsilon}{100}\) of \(R + \log 4\), and such that the angle between \(\tilde{\alpha}\) and \(\gamma\) is within \(\frac{\epsilon}{100}\)
of $\pi_2$. Here we assume that the two vectors (at the two diametrically opposite points) at $\gamma$ that are tangent to $\hat{\alpha}$ are on the same side of $\gamma$. Then to any such $\hat{\alpha}$ there is an $\varepsilon_{10}$-nearly homotopic $\alpha$ with endpoints on $\gamma$ (homotopic on $S$ through arcs with endpoints on $\gamma$) and such that $\alpha$ is orthogonal to $\gamma$. Each such $\alpha$ produces a pair of pants $\Pi \in \Pi_{\varepsilon,R}$ that contains $\gamma$ as a cuff.

Different $\alpha$’s give different pants. Two $\hat{\alpha}$ with the same $\alpha$ must have nearby endpoints along $\gamma$ and be $10\varepsilon$-nearly homotopic. So we get at least $\frac{2RN(S,\varepsilon)}{10\varepsilon}e^R$ of the $\alpha$’s and hence of the pants. □

Remark. Let $M > 1$, and let $X_\gamma(M)$ denote the number of pants in $\Pi_{\varepsilon,R}(\gamma)$ whose two other cuffs are in $\Gamma_{\varepsilon,R}$. Then $X_\gamma(M) \asymp Re^R$. The upper bound follows from the upper bound of the lemma. If the segment $\hat{\alpha}$ is of length within $\varepsilon_{100}$ of $2R - \text{hl}(\gamma) + \log 4$ and if the angle between $\hat{\alpha}$ and $\gamma$ is within $\varepsilon_{100}$ of $\frac{\pi}{2}$, then the induced pants have the desired property that the other two cuffs are in $\Gamma_{\varepsilon,R}$. On the other hand, it follows from the Connection Lemma that the number of such $\hat{\alpha}$’s is $\asymp N(S,\varepsilon,M)e^R$ for some constant $N(S,\varepsilon,M)$.

3.3. The Equidistribution Theorem. The following is the Equidistribution Theorem. This is a stronger equidistribution result than the one proved in our previous paper [4].

Theorem 3.1. Let $\varepsilon > 0$. Let $\mu$ be the measure on $\Pi_{\varepsilon,R}$ that assigns to each pants in $\Pi_{\varepsilon,R}(\gamma)$ the value 1. Then for $R$ large enough, the measure $\mu$ has the following properties:
(i) $\mu(\Pi_{\varepsilon,R}) \asymp e^{3R}$;
(ii) for every $\gamma \in \Gamma_{\varepsilon,R}$, the measure $\hat{\partial}_\pm \mu(\gamma)$ is $Ce^{-qR}$ equivalent to $K^\pm_\gamma \lambda_\pm(\gamma)$ or some constants $K^+_\gamma$ and $K^-_\gamma$ that satisfy the inequality
$$\left| \frac{K^+_\gamma}{K^-_\gamma} - 1 \right| < Ce^{-qR},$$
where $C, q > 0$ depend only on $S$ and $\varepsilon$;
(iii) moreover, $K^\pm_\gamma \asymp Re^R$ for $\gamma \in \Gamma_{\varepsilon,R}$.

Proof. The well-known result of Margulis [5] asserts that
$$\left| \Gamma_{\varepsilon,R} \right| \asymp \frac{e^{2R}}{R}.$$
the number of such pants that are to the right of $\gamma$. We first explain how to effectively estimate the two numbers.

Let $\eta_1$ and $\eta_2$ be two geodesic arcs on $S$ that connect the same two points on $S$. Denote by $[\cdot \eta_1 \cdot \eta_2 \cdot]$ the corresponding closed broken geodesic on $S$. (See the second paragraph of Section 4.1 for more on this notation.) We assume that $\eta_1$ and $\eta_2$ meet at the right angles and that we can orient $[\cdot \eta_1 \cdot \eta_2 \cdot]$ so that both right turns are “to the right” or both right turns are “to the left.”

Let $\gamma$ be the closed geodesic freely homotopic to $[\cdot \eta_1 \cdot \eta_2 \cdot]$. Then we can write

$$l(\gamma) = h(l(\eta_1), l(\eta_2))$$

for some smooth symmetric function $h$ (where $l$ denotes the length function), for which

$$h(e_1, e_2) = e_1 + e_2 - \log 2 + O(e^{-\min(e_1, e_2)/2}).$$

(The function $h$ can be computed explicitly from the basic formulas in hyperbolic geometry.)

Suppose that $\gamma$ is a (good) closed geodesic on $S$, and suppose that $\eta$ is a “third connection” for $\gamma$. So $\eta$ is a geodesic segment that meets $\gamma$ at two points $x$ and $y$, and is orthogonal to $\gamma$ at $x$ and $y$, and lies on the same side of $\gamma$ at both points: the two normal vectors to $\gamma$ pointing into $\eta$ at $x$ and $y$ are on the same component of $N^1(\gamma)$. Then there is a unique pair of pants $\Pi$ for which $\gamma \in \partial \Pi$ and for which $\eta$ is an orthogeodesic on $\Pi$ that lifts to be embedded in $\Pi$. Letting $\sigma_1$ and $\sigma_2$ be the two segments of $\gamma \setminus \{x, y\}$, we find that the two other cuffs $\gamma_1, \gamma_2$ of $\Pi$ are freely homotopic to $[\cdot \sigma_i \cdot \eta^{\pm 1} \cdot]$. So we have

$$l(\gamma_i) = h(l(\eta), l(\sigma_i)),$$

where $h$ is defined as above. Moreover, the two feet of $\Pi$ on $\gamma$ lie at the two midpoints of $x$ and $y$ on $\gamma$. (Really, we should think of $x$ and $y$ as lying on the parametrizing 1-torus for $\gamma$, and likewise the feet, but we will say that they are on $\gamma$, by a mild abuse of notation.)

Suppose that $I \subset N^1(\sqrt{\gamma})$ is an interval. So $I$ comprises a choice of component of $N^1(\gamma)$—a side of $\gamma$—along with a pair of intervals in $\gamma$, of equal length and placed halfway along $\gamma$ from each other. We should think of $\gamma$ as long, say longer than 10, and $I$ as short, say shorter than 1/10.

We define the region

$$\mathcal{R}(\gamma, I) \subset \gamma \times \gamma \times \mathbb{R}^+$$

as the set of $(x, y, l)$ for which

$$h(l, s_i) \in [2R - 2\varepsilon, 2R + 2\varepsilon]$$

and for which the two midpoints of $x$ and $y$ lie in the two intervals on $\gamma$ associated to $I$. Here $s_1, s_2$ are the lengths of the two arcs in $\gamma$ between $x$
and $y$. Then suppose $\eta$ is a third connection for $\gamma$ and that $\eta$ lies on the same side of $\gamma$ as $I$ does. Let $x$ and $y$ be the endpoints of $\eta$, and let $\Pi_\eta$ be the associated pair of pants for $\eta$ (and $\gamma$). Then $(x, y, l(\eta)) \in \mathcal{R}(\gamma, I)$ if and only if $\Pi_\eta$ is a good pair of pants, and the pair of feet of $\Pi_\eta$ on $N^1(\sqrt{\gamma})$ lies in $I$. Thus, the number of good pants whose feet belong to the interval $I$ is equal to the number of third connections $\eta$ (on a given side of $\gamma$) for which the associated triple $(x, y, l(\eta))$ lies in $\mathcal{R}(\gamma, I)$. So our goal is simply to count the number of third connections $\eta$ (on a given side of $\gamma$) for which the associated triple $(x, y, l(\eta))$ lies in $\mathcal{R}(\gamma, I)$.

One can see that the volume of $\mathcal{R}(\gamma, I)$ is on the order of $\varepsilon^2 |I|$ as follows. For any choice of a point in $I$, $x$ and $y$ are determined by $s_0$, and possible pairs $(l, s_0)$ lie in a diamond of size about $\varepsilon$. Since the area of the diamond is about $\varepsilon^2$, it follows that the volume of $\mathcal{R}(\gamma, I)$ is about $\varepsilon^2 |I|$.

If we denote by $C_\gamma$ the set of associated triples $(x, y, l(\eta))$ for all third connections $\eta$, our goal is simply to count $C_\gamma \cap \mathcal{R}(\gamma, I)$. The following counting formula is the main ingredient we need to finish the proof.

**Theorem 3.2.** There exist constants $C = C(S), q = q(S) > 0$ with the following properties. Let $\delta = e^{-qL}$, and suppose $a = b = \delta^2$. Then

$$
\#\text{Conn}_{A,B}(L_0, L_1) = \frac{1}{8\pi^2 \chi(S)} \frac{1}{\delta^2} e^{L} (1 + O(\delta)).
$$

This type of counting result appears often in literature (for example, see [3] [6]), and it goes back to Margulis [5].

Now let $Q$ be any cube of the form $J_1 \times J_2 \times [L_0, L_1] \subset \gamma \times \gamma \times \mathbb{R}^+$, with $|J_1| = |J_2| = L_1 - L_0 = \delta^2$ (and suppose $L_0, L_1$ are about $R$ that is large). Then Theorem 3.2 implies

$$
\#(Q \cap C_\gamma) = E S \delta^6 e^{L} (1 + O(\delta))
$$

$$
= (1 + O(\delta)) \int_Q E_s e^{L} dx dy dL,
$$
where

\[ E_S = \frac{1}{8\pi^8|\chi(S)|}. \]

It follows that for any region \( R \subset \gamma \times \gamma \times [R - 1, R + 1] \) tiled by such cubes,

\[ \#(R \cap C_\gamma) = (1 + O(\delta)) \int_R E_S e^L \, dx \, dy \, dL \]

and, more generally, for any region \( R \subset \gamma \times \gamma \times [R - 1, R + 1] \),

\[ \#(R \cap C_\gamma) = (1 + O(\delta)) \left( \int_R E_S e^L \, dx \, dy \, dL + \int_{N_{3\delta^2}(\partial R)} E_S e^L \, dx \, dy \, dL \right) , \]

where \( N_{3\delta^2}(\partial R) \) is the neighbourhood of \( \partial R \). (Here \( A \pm B \) means a number in \( [A-B, A+B] \).)

Now let \( R = R(\gamma, I) \) and assume that \( \varepsilon \geq |I| > \delta \). Then

\[ \Vol(N_{3\delta^2}(R)) \approx \varepsilon^2 \delta^2 = O(\delta \Vol(R)) \].

Therefore,

\[ \#(R \cap C_\gamma) = (1 + O(\delta)) \int_R E_S e^L \, dx \, dy \, dL \approx e^R \varepsilon^2 |I|. \]

On the other hand,

\[ \int_{R(\gamma, I)} E_S e^L \, dx \, dy \, dL = K_\gamma |I| \]

for some \( K_\gamma > 0 \) because the integral clearly depends only on \( |I| \) and

\[ \int_{R(\gamma, I)} \cdot = \int_{R(\gamma, I_1)} \cdot + \int_{R(\gamma, I_2)} \cdot , \]

where \( I_1, I_2 \) is a partition of \( I \).

Therefore,

\[ \#(R(\gamma, I) \cap C_\gamma) = K_\gamma |I| \left( 1 + O(\delta) \right) \]

for every interval \( I \) of length at least \( \delta \), and claim (ii) of the theorem follows. \( \Box \)

3.4. The Good Correction Theorem. To prove Theorem 2.3 we need to produce a measure \( \mu \) on good pants such that for each good geodesic \( \gamma \), \( \partial \mu(\gamma) \) is \( P(R) e^{-qR} \)-equivalent to some \( K_\gamma \lambda(\gamma) \), where \( \lambda \) is the Lebesgue measure on \( N^1(\gamma) \). In particular, \( \partial \mu(\gamma) \) must have the same total measure on both sides of \( \gamma \). In other words, there must be the same number of pants on both sides of \( \gamma \). We can write this as \( \partial \mu(\gamma) = 0 \).

Now let us construct the measure \( \mu_0 \) on good pants as the counting measure on the good pair of pants. Theorem 3.1 says that \( \partial \mu(\gamma) \) is \( Ce^{-qR} \)-equivalent to \( K^\pm \lambda_\pm(\gamma) \), with

\[ \left| \frac{K^+_\gamma}{K^-_\gamma} - 1 \right| < Ce^{-qR}. \]
So we have the desired equidistribution on each side of $\gamma$, and we have a small discrepancy between the number of pants on the two sides of $\gamma$. What we want to do is to make a small change in the number of each pair of good pants in order to correct the discrepancy. So we want to replace $\mu_0$ with $\mu_0 + X$ where $\partial X = -\partial \mu_0$ and $X$ is small compared to $\mu_0$.

To do this we consider the more general problem of finding $X$ such that

\[(2) \quad \partial X = \alpha,\]

and we ask two questions:

1. For which $\alpha$ can we solve (2)?
2. What bound can we get on the size of $X$ given a suitable bound on the size of $\alpha$?

It turns out that we can get fairly sharp answers for both questions. First, if $\partial X = \alpha$, then $\alpha \equiv 0$ in $H_1(S)$; we will prove that we can always solve $\partial X = \alpha$ when $\alpha$ is zero in singular homology. Second, if $\gamma$ is a single good closed geodesic and $\partial X = \gamma$, then

$$|X|(|\{\Pi : \gamma \in \partial \Pi\}| \geq 1,$$

and therefore

$$||X||_\infty \geq \frac{1}{Re^R}.$$

We prove that we can solve $\partial X = \alpha$ such that

$$||X||_\infty \leq P(R)e^{-R||\alpha||_\infty}$$

for some polynomial $P(R)$ depending on $S$ and $\varepsilon$. These two results are stated essentially as Theorems 3.4 and 3.5. The proof of these theorems will be the object of the remainder of this paper.

Having proven these theorems, we can correct the discrepancy and prove Theorem 2.3.

We let $\mu_1 = \mu_0 + X$, where $\partial X = -\partial \mu_0$, and

$$||X||_\infty \leq P(R)e^{-R||\partial \mu_0||_\infty} \leq P(R)e^{-R(1-q(S))R} \ll 1.$$

Then $\mu_1$ is a positive sum of good pants, has the same number of pants of both sides of every closed geodesic, and has the same equidistribution property as $\mu_0$ because it is a small perturbation. Therefore, it satisfies the conclusions of Theorem 2.3, and we can use it to build a good cover.

Recall that if $A$ is an Abelian group and $X$ any set, then $AX$ is the group of $A$-weighted finite formal sums of elements from $X$. 
Definition 3.2. Let $s_0, s_1 \in \mathbb{R} \Gamma_{\epsilon,R}$, and let $M \geq 1$. We say that $s_0 = s_1$ in $\Pi_{M\epsilon,R}$ homology if there exists $W \in \mathbb{R} \Pi_{M\epsilon,R}$ such that $\partial W = s_1 - s_0$.

The following theorem summarizes the main idea of this paper. It implies that every sum of good curves that is zero in the standard homology is the boundary of a sum of good pants. That is, if $s_0, s_1 \in \mathbb{R} \Gamma_{\epsilon,R}$ and $s_0 = s_1$ in the standard homology on the surface $S$, then $s_0 = s_1$ in $\Pi_{300\epsilon,R}$ homology. By $H_1(S)$ we denote the first homology on $S$ with rational coefficients.

**Theorem 3.3.** Let $\epsilon > 0$. There exists $R_0 = R_0(S, \epsilon) > 0$ such that for every $R > R_0$, the following holds. There exists a set $H = \{h_1, \ldots, h_{2g}\} \subset \mathbb{Q} \Gamma_{\epsilon,R}$ that form a basis of $H_1(S)$, such that for every $\gamma \in \Gamma_{\epsilon,R}$, there are $a_i \in \mathbb{Z}$ so that

$$\gamma = \sum_{i=1}^{2g} a_i h_i$$

in $\Pi_{300\epsilon,R}$ homology.

**Remark.** Observe that if $\gamma = 0$ in $H_1(S)$, then $\gamma = 0$ in $\Pi_{300\epsilon,R}$ homology.

The proof of this theorem occupies the primary text of Sections 4–9. But to prove the Ehrenpreis Conjecture we require the following stronger result.

**Theorem 3.4.** Let $\epsilon > 0$. There exists $R_0 > 0$ (which depend only on $S$ and $\epsilon$) such that for every $R > R_0$, there exists a set $H = \{h_1, \ldots, h_{2g}\} \subset \mathbb{Q} \Gamma_{\epsilon,R}$ and a map $\phi : \Gamma_{\epsilon,R} \to \mathbb{Q} \Pi_{300\epsilon,R}$ such that

(i) $h_1, \ldots, h_{2g}$ is a basis for $H_1(S)$;
(ii) $\partial(\phi(\gamma)) - \gamma \in \mathbb{Z}H$;
(iii) $\sum_{\gamma \in \Gamma_{\epsilon,R}} |\phi(\gamma)(\Pi)| < P(R)e^{-R}$ for every $\Pi \in \Pi_{300\epsilon,R}$, where the polynomial $P(R)$ depends only on $S$ and $\epsilon$.

**Remark.** Note that the map $\phi$ depends on $\epsilon$ and $R$, so sometimes we write $\phi = \phi_{\epsilon,R}$.

**Remark.** An inequality of the form

$$\sum_{\gamma \in \Gamma_{\epsilon,R}} |q(\gamma)(\Pi)| \leq A$$

is equivalent to saying

$$||q(\alpha)||_{\infty} \leq A||\alpha||_{\infty}$$

for all $\alpha \in \mathbb{Q} \Gamma_{\epsilon,R}$.

**Remark.** Observe that (i) and (ii) imply that $\tilde{\partial} \phi \tilde{\partial} = \tilde{\partial}$.

The existence of the function $\phi$ that satisfies conditions (i) and (ii) follows from Theorem 3.3. Condition (iii) will be proved using our randomization
As we go along, after every relevant homological statement we make randomization remarks. Theorem 3.4 then follows from these randomization remarks, as we explain at the end of Section 9.

Remark. Estimate (iii) in the statement of the theorem can be reformulated as follows. Consider the standard measures \( \sigma_\Gamma \) on \( \Gamma_{\varepsilon,R} \) and \( \sigma_\Pi \) on \( \Pi_{300\varepsilon,R} \). Then the map \( \phi \) is \( P(R) \)-semirandom with respect to \( \sigma_\Gamma \) and \( \sigma_\Pi \). See the appendix for definitions of the standard measures and the notion of semirandom maps.

The image of \( \phi \) lies in \( \Pi_{300\varepsilon,R} \), and we want it to lie in \( \Pi_{\varepsilon,R} \), so we require the following: Let \( M > 1 \). The following lemma states that any curve \( \gamma \in \Gamma_{\varepsilon,R} \) is homologous to some \( s \in R\Gamma_{\varepsilon,M,R} \) in \( \Pi_{\varepsilon,R} \) homology.

**Lemma 3.3.** Let \( \varepsilon, M > 0 \). Then there exists \( R_0 > 0 \) such that for every \( R > R_0 \), we can find a map \( q_M = q : \Gamma_{\varepsilon,R} \rightarrow Q^+\Pi_{\varepsilon,R} \) such that

(i) for every \( \gamma \in \Gamma_{\varepsilon,R} \), \( q(\gamma) \) is a positive sum of pants, all of which have \( \gamma \) as one boundary cuff (with the appropriate orientation), and two other cuffs in \( \Gamma_{\varepsilon,M,R} \), and \( \gamma - \partial q(\gamma) \in Q\Gamma_{\varepsilon,M,R} \);

(ii) for every \( \Pi \in \Pi_{\varepsilon,R} \), we have

\[
\sum_{\gamma \in \Gamma_{\varepsilon,R}} |q(\gamma)(\Pi)| \leq \frac{K}{R} e^{-R}
\]

for some constant \( K \equiv K(S, \varepsilon, M) > 0 \), where \( q(\gamma)(\Pi) \in Q^+ \) is the coefficient of \( \Pi \) in \( q(\gamma) \).

**Proof.** It follows from the remark after Lemma 3.2 that the number of pants that have \( \gamma \) as one boundary cuff and two other cuffs in \( \Gamma_{\varepsilon,M,R} \) is of the order \( Re^R \). Let \( q(\gamma) \) be the average of these pants. (The average of a finite set \( S \) is the formal sum of elements from \( S \) where each element in the sum has the weight \( \frac{1}{|S|} \).) The inequality in condition (ii) follows from the fact that for any \( \Pi \in \Pi_{1,R} \), the sum

\[
\sum_{\gamma \in \Gamma_{\varepsilon,R}} |q(\gamma)(\Pi)|,
\]

has at most three nonzero terms. \( \square \)

We can now state the following improved version of Theorem 3.4. This new theorem is exactly the same as the previous one except that the new function \( \phi \), which is denoted by \( \phi_{\text{new}} \), maps \( \Gamma_{\varepsilon,R} \) to \( Q\Pi_{\varepsilon,R} \) whereas the old \( \phi \) maps \( \Gamma_{\varepsilon,R} \) to \( Q\Pi_{300\varepsilon,R} \).

**Theorem 3.5.** Let \( \varepsilon > 0 \). There exists \( R_0 > 0 \) (which depend only on \( S \) and \( \varepsilon \)) such that for every \( R > R_0 \), there exist a set \( H = \{h_1, \ldots, h_{2g} \} \subset Q\Gamma_{\varepsilon,R} \) and a map \( \phi_{\text{new}} : \Gamma_{\varepsilon,R} \rightarrow Q\Pi_{\varepsilon,R} \) such that
(i) $h_1, \ldots, h_{2g}$ is a basis for $H_1(S)$;
(ii) $\partial(\phi_{\text{new}}(\gamma)) - \gamma \in \mathbb{Z}H$;
(iii) $\sum_{\gamma \in \Gamma_{\varepsilon,R}} |\phi_{\text{new}}(\gamma)(\Pi)| < P(R)e^{-R}$, where the polynomial $P(R)$ depends only on $S$ and $\varepsilon$.

Proof. We extend the function $\phi$ to $\mathbb{R}\Gamma_{\varepsilon,R}$ by linearity. For $\gamma \in \Gamma_{\varepsilon,R}$, we let
$$\phi_{\text{new}}(\gamma) = \phi(\gamma - \partial q(\gamma)) + q(\gamma),$$
where $\phi$ is from the previous theorem and $q = q_{300}$ is the improvement function from Lemma 3.3. Then since $H$ is a generating set for $H_1(S)$, it follows that $\partial(\phi_{\text{new}}(\gamma)) = \partial(\phi(\gamma))$, and thus we obtain $\partial(\phi_{\text{new}}(\gamma)) - \gamma \in \mathbb{Z}H$. It remains to verify inequality (iii) of the theorem.

For each $\Pi \in \Pi_{\varepsilon,R}$, we have
$$\sum_{\gamma} |\phi_{\text{new}}(\gamma)(\Pi)| \leq \sum_{\gamma} |q(\gamma)(\Pi)| + \sum_{\gamma} |\phi(\gamma - \partial q(\gamma))(\Pi)|. \tag{3}$$

On the other hand, for each $\eta \in \Gamma_{\varepsilon,R}$, we have the inequality
$$\sum_{\gamma} |\partial q(\gamma)(\eta)| \leq C_1$$
for some universal constant $C_1 > 0$. In other words, the total weight of $\eta$ in the formal sum of curves
$$\sum_{\gamma} \partial q(\gamma) \in Q\Gamma_{\varepsilon,R}$$
is at most $C_1$. This follows from the last inequality of Lemmas 3.3 and 3.2.

Thus, we have
$$\sum_{\gamma} |\phi(\gamma - \partial q(\gamma))(\Pi)| \leq \sum_{\gamma} |\phi(\gamma)(\Pi)| + C_1 \sum_{\gamma} |\phi(\gamma)(\Pi)|.$$

We replace this inequality in (3), and the theorem follows.

Of course $\phi_{\text{new}}$ extends linearly to $Q\Gamma_{\varepsilon,R}$. We observe that if $\gamma$ is zero in $H_1(S)$, then $\partial\phi_{\text{new}}(\gamma)$ is equal to $\gamma$ because $\partial(\phi_{\text{new}}(\gamma)) - \gamma \in \mathbb{Z}H$ and $\partial(\phi_{\text{new}}(\gamma))$ differs from $\gamma$ by a boundary. In particular, for any $\mu \in Q\Pi_{\varepsilon,R}$, the equality
$$\partial\phi_{\text{new}}\partial\mu = \partial\mu$$
holds.

3.5. A proof of Theorem 2.3. First we state and prove the following lemma about equivalent measures on the circle $\mathbb{R}/2\mathbb{R}$, where $R > 0$ is a parameter. Recall that $\lambda$ denotes the Lebesgue measure on the circle $\mathbb{R}/2\mathbb{R}$.

Lemma 3.4. If $\alpha$ is $\delta$-equivalent to $K\lambda$ on $\mathbb{R}/2\mathbb{R}$, for some $K > 0$, and $\beta$ is a measure on $\mathbb{R}/2\mathbb{R}$, then $\alpha + \beta$ is $(\frac{|\beta|}{2R} + \delta)$-equivalent to $(K + \frac{|\beta|}{2R})\lambda$ on $\mathbb{R}/2\mathbb{R}$. \hfill $\square$
Proof. Recall the definition of $\delta$-equivalent measures from the previous section. We need to prove that $(\alpha + \beta)(I) \leq (K + \frac{|\beta|}{2R})|N_{\delta'}(I)|$, where $\delta' = \delta + \frac{|\beta|}{2R}$, for any interval $I$ such that $\delta'$ neighbourhood of $I$ is a proper subset of the circle $\mathbb{R}/2\mathbb{Z}$.

We have

\[(\alpha + \beta)(I) \leq K(|I| + 2\delta) + |\beta| \leq \left(K + \frac{|\beta|}{2R}\right)(|I| + 2\delta + \frac{|\beta|}{K}) = \left(K + \frac{|\beta|}{2R}\right)|N_{\delta'}(I)|.\]

\[
\square
\]

We give the following definitions. To any measure $\alpha \in M(N^1(\sqrt{\gamma}))$ we associate the number $|\alpha(\gamma)| = |\alpha_+(\gamma)| + |\alpha_-(\gamma)|$.

We proceed with the proof of Theorem 2.3. Let $\mu$ be the measure on $\Pi_{\varepsilon,R}$ from Theorem 3.1. Define the measure $\mu_1$ on $\Pi_{\varepsilon,R}$ by letting $\mu_1 = \mu - \phi(\partial \mu)$, where $\phi = \phi_{\text{new}}$ is from Theorem 3.5. We show that $\mu_1$ is the measure that satisfies the conclusions of Theorem 2.3.

As observed before, $\partial \mu_1 = 0$. It remains to show that $\hat{\partial} \mu_1(\gamma)$ is $P(R)e^{-qR}$-equivalent to some multiple of the Lebesgue measure on $N^1(\sqrt{\gamma})$. Recall from Theorem 3.1 that the measure $\hat{\partial} \mu(\gamma)$ is $Ce^{-qR}$ equivalent to $K_+^\pm \lambda_\pm(\gamma)$ for some constants $K_+^+$ and $K_-$ that satisfy the inequality

\[
\frac{K_+^+}{K_-^+} - 1 < Ce^{-qR},
\]

and recall that $K_+^+ \asymp Re^R$. Then for all $\gamma$, from (4) we get

\[
|\partial \mu(\gamma)| \leq CR e^{(1-q)R},
\]

which together with Theorem 3.5 yields the inequality

\[
|\phi(\partial \mu)(\Pi)| \leq P(R)e^{-qR}
\]

for each pair of pants $\Pi$, and we obtain

\[
|\hat{\partial} \phi(\partial \mu)(\gamma)| \leq P(R)e^{(1-q)R}
\]

for all $\gamma$. Since $K_+^+ \asymp Re^R$, we conclude from Lemma 3.4 that $\hat{\partial} \mu_1(\gamma)$ is $P(R)e^{-qR}$-equivalent to some multiple of the Lebesgue measure $\lambda(\gamma)$ on $N^1(\sqrt{\gamma})$, and we are done.

4. The theory of inefficiency

In this section we develop the theory of inefficiency. This theory supports the geometric side of the correction theory that is used to prove our main technical result the Good Correction Theorem.
Before we begin with the estimates of this section, we will provide a brief overview of the remainder of this paper. Our goal is to prove identities in the good pants homology, which means that we need to generate a formal sum of good pants whose boundary is a certain given formal sum of good curves. How do we generate a pair of good pants?

We generate a pair of good pants by constructing a Θ-graph made out of geodesic segments; the Connection Lemma insures that we have enough geodesic segments with the desired properties, and the Theory of Inefficiency allows us to estimate the length of the cuffs of the associated pair of pants in terms of the length of the geodesic segments and the angle at which they meet.

In every identity we prove in the good pants homology, we will state our hypothesis in terms of the Theory of Inefficiency, and every time we estimate the length of a geodesic segment or a closed geodesic, we will use this theory as well.

4.1. The inefficiency of a piecewise geodesic arc. By $T^1\mathbb{H}^2$ we denote the unit tangent bundle of $\mathbb{H}^2$. Elements of $T^1\mathbb{H}^2$ are pairs $(p,u)$, where $p \in \mathbb{H}^2$ and $u \in T^1_p\mathbb{H}^2$. For $u,v \in T^1_p\mathbb{H}^2$, we let $\Theta(u,v)$ denote the unoriented angle between $u$ and $v$. Let $\alpha : [a,b] \to \mathbb{H}^2$ be a unit speed geodesic segment. We let $i(\alpha) = \alpha'(a)$, and $t(\alpha) = \alpha'(b)$.

Let $\alpha_1, \ldots, \alpha_n$ denote oriented piecewise geodesic arcs on $S$ such that the terminal point of $\alpha_i$ is the initial point of $\alpha_{i+1}$. By $\alpha_1 \alpha_2 \ldots \alpha_n$ we denote the concatenation of the arcs $\alpha_i$. If the initial point of $\alpha_1$ and the terminal point of $\alpha_n$ are the same, by $[\alpha_1 \alpha_2 \ldots \alpha_n]$ we denote the corresponding closed curve.

We define the inefficiency operator as follows. We first discuss the inefficiency of piecewise geodesic arcs and after that the inefficiency of piecewise geodesic closed curves.

Definition 4.1. Let $\alpha$ be an arc on a surface. By $\gamma$ we denote the geodesic arc with the same endpoints and homotopic to $\alpha$. We let $I(\alpha) = I(\alpha) - I(\gamma)$. We call $I(\alpha)$ the inefficiency of $\alpha$. (The inefficiency $I(\alpha)$ is equal to 0 if and only if $\alpha$ is a geodesic arc.)

We observe the monotonicity of inefficiency. Let $\alpha, \beta, \gamma \subset \mathbb{H}^2$ be three piecewise geodesic arcs in $\mathbb{H}^2$ such that $\alpha \beta \gamma$ is a well-defined piecewise geodesic arc. Then $I(\alpha \beta \gamma) \geq I(\beta)$. This is seen as follows. Let $\eta$ be the geodesic arc with the same endpoints as $\alpha \beta \gamma$, and let $\beta'$ be the geodesic arc with the same endpoints as $\beta$. Then

\[
I(\alpha \beta \gamma) = I(\alpha \beta \gamma) - I(\eta) \\
\geq I(\alpha \beta \gamma) - I(\alpha \beta') \\
= I(\beta) - I(\beta') \\
= I(\beta).
\]
We also define the inefficiency function of an angle $\theta \in [0, \pi]$ as follows. Let $\alpha_\infty$ and $\beta_\infty$ be two geodesic rays in $\mathbb{H}^2$ that have the same initial point, such that $\theta$ is the exterior angle between $\alpha_\infty$ and $\beta_\infty$. (Then $\theta$ is also the bending angle of the piecewise geodesic $\alpha_\infty^{-1}\beta_\infty$.) We want to define $I(\theta)$ as the inefficiency of $\alpha_\infty^{-1}\beta_\infty$, but since the piecewise geodesic $\alpha_\infty^{-1}\beta_\infty$ is infinite in length, we need to prove that such a definition is valid.

Consider a geodesic triangle in $\mathbb{H}^2$ with sides $A, B$ and $C$, and let $\theta > 0$ be the exterior angle opposite to $C$. (We also let $l(A) = A, l(B) = B, l(C) = C$.) Then

$$\cosh C = \cosh A \cosh B + \cos \theta \sinh A \sinh B$$

and, therefore,

$$\cosh C = \cosh A \cosh B + \cos \theta \sinh A \sinh B$$

$$\frac{\cosh C}{e^{A+B}} = \frac{\cosh A \cosh B}{e^A} - \frac{\cos \theta \sinh A \sinh B}{e^A}.$$ 

We conclude that

$$\frac{\cosh C}{e^{A+B}} \rightarrow \frac{1}{4} (\cos \theta + 1),$$

when $A, B \rightarrow \infty$. Since

$$\frac{\cosh C}{e^C} \rightarrow \frac{1}{2}, C \rightarrow \infty,$$

we get

$$e^{C-A-B} \rightarrow \left( \cos \frac{\theta}{2} \right)^2,$$

and therefore

$$A + B - C \rightarrow 2 \log \sec \frac{\theta}{2}.$$ 

Let $r, s > 0$, and let $\alpha_r$ be the geodesic subsegment of $\alpha_\infty$ (with the same initial point) of length $r$. Similarly, $\beta_s$ is a geodesic subsegment of $\beta_\infty$ of length $s$. Then we let

$$I(\theta) = I(\alpha_\infty^{-1}\beta_\infty) = \lim_{r, s \rightarrow \infty} I(\alpha_r^{-1}\beta_s).$$

It follows from the above discussion that this limit exists and

$$I(\theta) = 2 \log \sec \frac{\theta}{2}.$$ (5)

4.2. Preliminary lemmas. We have the following lemma.

**Lemma 4.1.** Let $\alpha$ denote an arc on $S$, and let $\gamma$ be the appropriate geodesic arc with the same endpoints as $\alpha$ and homotopic to $\alpha$. Choose lifts of $\alpha$ and $\gamma$ in the universal cover $\mathbb{H}^2$ that have the same endpoints, and let $\pi : \alpha \rightarrow \gamma$ be the nearest point projection. Let

$$E(\alpha) = \sup_{x \in \alpha} d(x, \pi(x)).$$

Then

$$E(\alpha) \leq \frac{I(\alpha)}{2} + \log 2.$$
Figure 1. The case where $\gamma$ is an arc

Proof. Let $E > 0$. The minimally inefficient arc $\alpha$ (which has the same endpoints as $\gamma$ and is homotopic to $\gamma$) that is at the distance $E$ from $\gamma$ is given in Figure 7. Here $\gamma$ is divided into two subsegments of length $L^-$ and $L^+$. Let $A^- = I(\alpha^-)$ and $A^+ = I(\alpha^+)$. By the monotonicity of inefficiency, and using the inefficiency for angles, we have

$$E + L^- - A^- < I\left(\frac{\pi}{2}\right)$$

and

$$E + L^+ - A^+ < I\left(\frac{\pi}{2}\right).$$

Summing up yields

$$E < \frac{I(\alpha)}{2} + I\left(\frac{\pi}{2}\right) = I(\alpha) + \log 2,$$

since by (5), we have $I(\frac{\pi}{2}) = \log 2$. \hfill \Box

The following is the New Angle Lemma.

Lemma 4.2 (New Angle Lemma). Let $\delta, \Delta > 0$, and let $\alpha\beta \subset \mathbb{H}^2$ be a piecewise geodesic arc, where $\alpha$ is piecewise geodesic arc and $\beta$ is a geodesic arc. Suppose that $\gamma$ is the geodesic arc with the same endpoints as $\alpha\beta$. There exists $L = L(\delta, \Delta) > 0$ such that if $I(\beta) > L$ and $I(\alpha\beta) \leq \Delta$, then the unoriented angle between $\gamma$ and $\beta$ is at most $\delta$.

Proof. Denote by $\theta$ the angle between $\gamma$ and $\beta$, and let $h$ be the distance between the other endpoint of $\beta$ and $\gamma$. Then

$$\frac{\sinh(h)}{\sin(\theta)} = \sinh(I(\beta)).$$

The lemma follows from this equation and the fact that $h$ is bounded in terms of $I(\alpha\beta)$ (see Lemma 4.1). \hfill \Box
We also have

**Lemma 4.3.** Suppose that $\alpha\beta\gamma$ is a concatenation of three geodesic arcs in $\mathbb{H}^2$, and let $\theta_{\alpha\beta}$ and $\theta_{\beta\gamma}$ be the two bending angles. Suppose that $\theta_{\alpha\beta}, \theta_{\beta\gamma} < \frac{\pi}{2}$. Then

$$I(\alpha\beta\gamma) \leq \log(\sec(\theta_{\alpha\beta})) + \log(\sec(\theta_{\beta\gamma})).$$

**Proof.** Let $\eta_1$ be the geodesic that is orthogonal to $\beta$ at the point where $\alpha$ and $\beta$ meet. Similarly, let $\eta_2$ be the geodesic that is orthogonal to $\beta$ at the point where $\beta$ and $\gamma$ meet. Let $A_\alpha$ be the geodesic arc orthogonal to $\eta_1$ that starts at the initial point of $\alpha$, and let $A_\gamma$ be the geodesic arc orthogonal to $\eta_2$ that starts at the terminal point of $\gamma$.

Let $\eta$ be the geodesic arc with the same endpoints as $\alpha\beta\gamma$. Then $I(\eta) \geq I(A_\alpha) + I(\beta) + I(A_\gamma)$.

On the other hand, from the hyperbolic law of sines, we have

$$\sinh A_\alpha = \sinh \alpha \cdot \cos \theta_{\alpha\beta},$$

and hence

$$\log \sinh \alpha - \log \sinh A_\alpha = \log \sec \theta_{\alpha\beta},$$

which implies

$$\alpha - A_\alpha \leq \log \sec \theta_{\alpha\beta},$$

because the derivative of $\log \sinh$ is greater than one. Thus, we have proved that $I(A_\alpha) > I(\alpha) - \log(\sec(\theta_{\alpha\beta}))$, and similarly $I(A_\gamma) > I(\gamma) - \log(\sec(\theta_{\beta\gamma}))$. So

$$I(\alpha\beta\gamma) < I(\alpha) + I(\gamma) - I(A_\alpha) - I(A_\gamma)$$

$$< \log(\sec(\theta_{\alpha\beta})) + \log(\sec(\theta_{\beta\gamma})).$$

$\square$

4.3. The Long Segment Lemmas for arcs. We now state and prove several lemmas called the Long Segment Lemmas. The following is the Long Segment Lemma for angles.

**Lemma 4.4 (Long Segment Lemma).** Let $\delta > 0, \Delta > 0$. There exists a constant $L = L(\delta, \Delta) > 0$ such that if $\alpha$ and $\beta$ are oriented geodesic arcs such that $I(\alpha\beta) \leq \Delta$ (assuming that the terminal point of $\alpha$ is the initial point of $\beta$) and $I(\alpha), I(\beta) > L$, then $I(\alpha\beta) < I(\Theta(t(\alpha), i(\beta))) < I(\alpha\beta) + \delta$.

**Proof.** The left hand-side of the above inequality follows from the monotonicity of inefficiency. For the right-hand side, let $\alpha_\infty$ and $\beta_\infty$ denote the geodesic rays whose initial point is the point where $\alpha$ and $\beta$ meet and that contain $\alpha$ and $\beta$ respectively. (We also assume that $\alpha_\infty$ has the same orientation as $\alpha$ and $\beta_\infty$ the same orientation as $\beta$.) Recall that $I(\alpha_\infty\beta_\infty)$ was defined just above formula (5). Let $\eta$ be the geodesic arc with the same endpoints as $\alpha\beta$, and let $\eta_1$ be the geodesic ray with the same endpoints as $\alpha_\infty\beta_\infty$. By $\theta_0$ we
denote the angle between $\alpha$ and $\eta$, by $\theta_1$ the angle between $\eta$ and $\beta$, and by $\theta_2$ the angle between $\eta$ and $\eta_1$.

We observe that $\theta_0$, $\theta_1$, and $\theta_2$ are small (by the New Angle Lemma), and therefore

$$I(\alpha \infty \beta \infty) < I(\alpha \infty \beta) + I(\theta_0 + \theta_2)$$

$$< I(\alpha \beta) + I(\theta_0 + \theta_2)$$

$$< I(\alpha \beta) + o(1),$$

where $o(1) \to 0$ as $L \to \infty$.

The following is the Long Segment Lemma for arcs.

**Lemma 4.5** (Long Segment Lemma for arcs). Suppose we can write $\eta = \alpha \beta \gamma$, where $\alpha$ and $\gamma$ are piecewise geodesic arcs and $\beta$ is a geodesic arc of length $l$. Then

$$\left| I(\alpha \beta) + I(\beta \gamma) - I(\alpha \beta \gamma) \right| \to 0$$

uniformly when $l \to \infty$ and $I(\alpha \beta) + I(\beta \gamma)$ is bounded above.

**Proof.** If we replace $\alpha$ and $\gamma$ by the associated geodesics arcs, then $I(\alpha \beta) + I(\beta \gamma) - I(\alpha \beta \gamma)$ will be unchanged, and $I(\alpha \beta) + I(\beta \gamma)$ will be decreased, so we can assume that $\alpha$ and $\gamma$ are geodesic arcs. We divide $\beta$ at its midpoint into $\beta^-$ and $\beta^+$, so $\beta = \beta^- \beta^+$, and $\alpha \beta \gamma = \alpha \beta^- \beta^+ \gamma$. We will show the following estimates (for $\delta$ small when $l$ is large and $I(\alpha \beta) + I(\beta \gamma)$ is bounded above):

(i) $$\left| I(\alpha \beta^-) + I(\beta^+ \gamma) - I(\alpha \beta \gamma) \right| < \delta,$$

(ii) $$\left| I(\alpha \beta) + I(\alpha \beta^-) \right| < \delta,$$

(iii) $$\left| I(\beta \gamma) + I(\beta^+ \gamma) \right| < \delta.$$

The lemma then follows from (i), (ii), and (iii).

For (i), we refer to Figure 2. We find that

$$0 \leq I(\hat{\alpha} \hat{\beta}) = I(\alpha \beta \gamma) - I(\alpha \beta^-) - I(\beta^+ \gamma).$$

![Figure 2. The Long Segment Lemma](image)
Moreover, when $I(\alpha \beta^-)$ (which is at most $I(\alpha \beta)$) and $I(\beta^-)$ (which is at most $I(\beta \gamma)$) are bounded above, and $I(\beta^-) = I(\beta^+)$ is a large, then $\theta^-$ and $\theta^+$ are small (by the New Angle Lemma), so $I(\hat{\alpha} \hat{\beta}) \leq I(\theta^- + \theta^+)$ is small. Likewise, $I(\alpha \beta) - I(\alpha \beta^-) = I(\hat{\alpha} \hat{\beta}^+)$ and $0 \leq I(\hat{\alpha} \hat{\beta}^+) \leq I(\theta^-)$. This proves (i) and (ii), and (iii) is the same as (ii).

4.4. The inefficiency of a closed piecewise geodesic curve. Let $\alpha_1, \ldots, \alpha_n$ denote oriented piecewise geodesic arcs on $S$ such that the terminal point of $\alpha_i$ is the initial point of $\alpha_{i+1}$. By $\alpha_1 \alpha_2 \ldots \alpha_n$ we denote the concatenation of the arcs $\alpha_i$. Assume that the initial point of $\alpha_1$ and the terminal point of $\alpha_n$ are the same. By $[\alpha_1 \alpha_2 \ldots \alpha_n]$ we denote the corresponding closed curve.

We define the inefficiency operator as follows.

Definition 4.2. Let $\alpha$ be a closed curve on a surface. By $\gamma$ we denote the appropriate closed geodesic that is freely homotopic to $\alpha$. We let $I(\alpha) = l(\alpha) - l(\gamma)$. We call $I(\alpha)$ the inefficiency of $\alpha$. (The inefficiency $I(\alpha)$ is equal to 0 if and only if $\alpha$ is a closed geodesic.)

The following is a closed curve version of Lemma 4.1.

Lemma 4.6. Let $\alpha$ denote a closed curve on $S$, and let $\gamma$ be the appropriate closed geodesic freely homotopic to $\alpha$. Choose lifts $\tilde{\alpha}$ and $\tilde{\gamma}$, of $\alpha$ and $\gamma$ respectively, in the universal cover $\mathbb{H}^2$ that have the same endpoints. The nearest point projection $\tilde{\pi} : \tilde{\alpha} \to \tilde{\gamma}$ descends to the map $\pi : \alpha \to \gamma$. Let

$$E = E(\alpha) = \sup_{x \in \tilde{\alpha}} d(x, \tilde{\pi}(x)).$$

Then providing $I(\alpha) > L_0$ for some universal constant $L_0$, we have

$$E \leq \frac{I(\alpha)}{2} + 1.
$$

Proof. The minimally inefficient closed curve $\alpha$ that is freely homotopic to $\gamma$ and the distance $E$ from $\gamma$ is given in Figure 3. Denote by $\eta$ the corresponding geodesic segment of length $E$. Then by the Long Segment Lemma and monotonicity of inefficiency,

$$I(\eta \gamma \eta^{-1}) \leq I(\eta \gamma) + I(\gamma \eta^{-1}) + \frac{1}{7}\pi,$$

$$\leq I\left(\frac{\pi}{2}\right) + I\left(\frac{\pi}{2}\right) + \frac{1}{7}\pi,$$

$$< 2,$$

providing that $I(\gamma) > L_0$, where $L_0$ is a universal constant. Hence

$$I(\alpha) - I(\gamma) \geq 2I(\eta) - 2,$$

or

$$E \leq \frac{I(\alpha)}{2} + 1. \quad \Box$$
The following is the Long Segment Lemma for closed curves.

**Lemma 4.7** (Long Segment Lemma for closed curves). Let $\alpha$ be an piecewise geodesic arc and $\beta$ a geodesic arc on $S$, such that the initial point of $\alpha$ is the terminal point of $\beta$ and the initial point of $\beta$ is the terminal point of $\alpha$. Then

$$|I([\alpha \beta]) - I(\beta \alpha \beta)| < \delta,$$

where $\delta \to 0$ when $l(\beta) \to \infty$ and $I(\beta \alpha \beta)$ is bounded above.

*Proof*. The proof is similar to the proof of Lemma 4.5 and is left to reader. \qed

4.5. The Sum of Inefficiencies Lemma. The following is the Sum of Inefficiencies Lemma. Let $S$ denote a closed hyperbolic Riemann surface

**Lemma 4.8** (Sum of Inefficiencies Lemma). Let $\varepsilon, \Delta > 0$ and $n \in \mathbb{N}$. There exists $L = L(\varepsilon, \Delta, n) > 0$ such that the following holds. Let $\alpha_1, \ldots, \alpha_{n+1} = \alpha_1, \beta_1, \ldots, \beta_n$ be geodesic arcs on the surface $S$ such that $\alpha_1 \beta_1 \alpha_2 \beta_2 \ldots \alpha_n \beta_n$ is a piecewise geodesic closed curve on $S$. If $I(\alpha_i \beta_i \alpha_{i+1}) \leq \Delta$ and $l(\alpha_i) \geq L$, then

$$\left| I([\alpha_1 \beta_1 \alpha_2 \beta_2 \ldots \alpha_n \beta_n]) - \sum_{i=1}^{n} I(\alpha_i \beta_i \alpha_{i+1}) \right| \leq \varepsilon.$$

*Proof*. It directly follows from the Long Segment Lemma for closed curves. \qed
Remark. In particular, we can leave out the $\beta$’s in the above lemma and write

$$\left|I([\alpha_1\alpha_2\ldots\alpha_n]) - \sum_{i=1}^{n} I(\alpha_i\alpha_{i+1})\right| \leq \varepsilon,$$

providing that $I(\alpha_i\alpha_{i+1}) \leq \Delta$ and $I(\alpha_i) \geq L$. Moreover, by the Long Segment Lemma for Angles (for $L$ large enough), we have

$$\left|I([\alpha_1\alpha_2\ldots\alpha_n]) - \sum_{i=1}^{n} I(\theta_i)\right| \leq 2\varepsilon,$$

where $\theta_i = \Theta(t(\alpha_i), i(\alpha_{i+1}))$.

A more general version of the Sum of Inefficiencies Lemma is as follows. (The proof is the same.)

**Lemma 4.9.** Let $\varepsilon, \Delta > 0$ and $n \in \mathbb{N}$. There exists $L = L(\varepsilon, \Delta, n) > 0$ such that the following holds. Let

$$\alpha_1, \ldots, \alpha_{n+1} = \alpha_1$$

and

$$\beta_{11}, \ldots, \beta_{1j_1}, \ldots, \beta_{n1}, \ldots, \beta_{nj_n}$$

be geodesic segments on $S$ such that $\alpha_1\beta_{11}\ldots\beta_{1j_1}\ldots\alpha_n\beta_{n1}\ldots\beta_{nj_n}$ is a piecewise geodesic closed curve on $S$. If $I(\alpha_i\beta_i\alpha_{i+1}) \leq \Delta$ and $I(\alpha_i) \geq L$, then

$$\left|I([\alpha_1\beta_{11}\ldots\beta_{1j_1}\ldots\alpha_n\beta_{n1}\ldots\beta_{nj_n}]) - \sum_{i=1}^{n} I(\alpha_i\beta_i)\right| \leq \varepsilon.$$

**Proof.** It directly follows from the Long Segment Lemma. \qed

We will use the theory of inefficiency for two purposes: first, to control the geometry of piecewise geodesic arcs and closed curves, and second, to precisely estimate the length of the associated geodesic arcs and closed curves.

For example, suppose that $[\alpha, \beta]$ is a closed curve, that $\alpha$ and $\beta$ meet nearly at right angles at the two places they meet, and that $\alpha$ and $\beta$ are both long. Then $I(\gamma)$ is close to $I(\alpha) + I(\beta) - \log 4$ where $\gamma$ is the corresponding closed geodesic.

As a second example, suppose $\alpha_1, \alpha_2, \alpha_3$ is a piecewise geodesic arc, and let $\alpha_{12}$ be the geodesic arc homotopic rel endpoints to $\alpha_1\alpha_2$. Likewise, define $\alpha_{23}$ and $\alpha_{123}$. Then

$$I(\alpha_{123}) = I(\alpha_1) + I(\alpha_2) + I(\alpha_3) + I(\alpha_{12}) + I(\alpha_{23}),$$

provided that $I(\alpha_{12}), I(\alpha_{23}) \leq \Delta$ and $I(\alpha_2) \geq L(\Delta)$. 
5. The Geometric Square Lemma

In this section we will prove the Geometric Square Lemma (Lemma 5.4), which will then be reformulated as the Algebraic Square Lemma in the next section. This section is probably the hardest and most technical section in the paper, and the reader who is still struggling to have a clear idea of where we are going may wish to skip to the next section and see how the Algebraic Square Lemma follows from the Geometric one.

From now on we can think of the surface $S$ as being fixed. We also fix $\varepsilon > 0$. However, for the reader’s convenience we always emphasize how quantities may depend on $S$ and $\varepsilon$.

5.1. Notation and preliminary lemmas. By an oriented closed geodesic $C$ on $S$ we will mean an isometric immersion $C : \mathbb{T}_C \rightarrow S$, where $\mathbb{T}_C = \mathbb{R}/l(C)$, and $l(C)$ is the length of $C$. To simplify the notation, by $C : \mathbb{R} \rightarrow S$ we also denote the corresponding lift. (Such a lift is uniquely determined once we fix a covering map $\pi : \mathbb{R} \rightarrow \mathbb{T}_C$.) We call $\mathbb{T}_C$ the parametrizing torus for $C$ (because $\mathbb{T}_C = \mathbb{R}/l(C)$ is a 1-torus). By a point on $C$ we mean $C(p)$ where $p \in \mathbb{T}_C$, or $p \in \mathbb{R}$. Given two points $a, b \in \mathbb{R}$, we let $C[a, b]$ be the restriction of $C : \mathbb{R} \rightarrow S$ to the interval $[a, b]$. If $b < a$, then the orientation of the segment $C[a, b]$ is the negative of the orientation for $C$. Of course $C[a + nl(C), b + nl(C)]$ is the same creature for $n \in \mathbb{Z}$. By $C'(p)$ we denote the unit tangent vector to $C$ with the appropriate orientation.

Recall that $T^1 \mathbb{H}^2$ denotes the unit tangent bundle, where elements of $T^1(\mathbb{H}^2)$ are pairs $(p, u)$, where $p \in \mathbb{H}^2$ and $u \in T^1_p \mathbb{H}^2$. The tangent space $T^1_p$ has a complex structure, and given $u \in T^1_p \mathbb{H}^2$, by $\sqrt{-1}u \in T^1_p \mathbb{H}^2$ we denote the vector obtained from $u$ by rotating for $\pi/2$.

Recall that for $u, v \in T^1_p \mathbb{H}^2$, we let $\Theta(u, v)$ denote the unoriented angle between $u$ and $v$. If $u \in T^1_p \mathbb{H}^2$, then $u @ q \in T^1_q \mathbb{H}^2$ denotes the vector $u$ parallel transported to $q$ along the geodesic segment connecting $p$ and $q$. We use similar notation for points in $T^1 \mathbb{H}^2$, except that in this case one always has to specify the segment between $p$ and $q$ along which we parallel transport vectors from $T^1_p \mathbb{H}^2$ to $T^1_q \mathbb{H}^2$.

We refer to the following lemma as the Convergence Lemma. The proof is left to the reader.

**Lemma 5.1.** Suppose $A$ and $B$ are oriented geodesics in $\mathbb{H}^2$ that are $E$-nearly homotopic, and let

$$a : \left[ -\frac{l(A)}{2}, \frac{l(A)}{2} \right] \rightarrow \mathbb{H}^2, \quad b : \left[ -\frac{l(B)}{2}, \frac{l(B)}{2} \right] \rightarrow \mathbb{H}^2$$

denote the unit time parametrization. Set $l = \frac{1}{2} \min(l(A), l(B))$. Then there exists $0 \leq t_0 \leq E$ such that for $t \in [-l, l]$, the following inequalities hold:
We define the distance function
\[ \text{dis}((p, u), (q, v)) = \max(\Theta(u @ q, v), d(p, q)). \]
(We do not insist that dis is a metric on \( T^1(\mathbb{H}^2) \).

Lemma 5.2. Let \( \varepsilon, L > 0 \). There exists a constant \( \varepsilon'(L) \) with the following properties. Suppose that \( \alpha : [a_0, a_1] \to \mathbb{H}^2 \) and \( \beta : [b_0, b_1] \to \mathbb{H}^2 \) are \( \varepsilon \)-nearly homotopic, that is, \( d(\alpha(a_i), \beta(b_i)) \leq \varepsilon \). Suppose that \( a_1 - a_0 > L \), and \( \varepsilon < 1 \). Then
\[ \text{dis}(\alpha'(a_i), \beta'(b_i)) \leq \varepsilon(1 + \varepsilon'(L)), \]
with \( \varepsilon'(L) \to 0 \) as \( L \to \infty \).

5.2. The Preliminary Geometric Square Lemma (the PGSL). Suppose
\( C_{ij} \), \( i, j = 0, 1 \), are four closed geodesics on \( S \), and imagine that \( C_{ij} \) is covered by two overlapping arcs \( C_{ij}^+ \) and \( C_{ij}^- \), where \( C_{i0}^+ \) and \( C_{i0}^- \) are nearly homotopic and likewise for \( C_{0j}^+ \) and \( C_{1j}^- \). The Geometric Square Lemma (GSL) states that
\[ \sum (-1)^i C_{ij} = 0. \]
The full statement of the GSL is given in Section 5.3.

The following is the Preliminary Geometric Square Lemma. We have
added hypothesis (v) to the GSL (Lemma 5.4) so as to find points in the two convergence intervals of the four curves, which are nearly diametrically opposite.

Lemma 5.3 (Preliminary Geometric Square Lemma). Let \( E, \varepsilon > 0 \). There exist constants \( K = K(\varepsilon, E) > 0 \) and \( R_0(S, \varepsilon, E) > 0 \) with the following properties. Suppose that we are given four oriented geodesics \( C_{ij} \in \Gamma_{\varepsilon,R}, i, j = 0, 1 \), and for each \( ij \) we are given four real numbers \( x_{ij}^- < x_{ij}^+ < y_{ij}^- < y_{ij}^+ < x_{ij}^+ + 1(C_{ij}). \) Assume that
(i) The inequalities \( x_{ij}^+ - x_{ij}^- > K \) and \( y_{ij}^+ - y_{ij}^- > K \), hold.
(ii) the segments \( C_{ij}[x_{ij}^-, x_{ij}^+] \) and \( C_{i'j'}[x_{i'j'}^-, x_{i'j'}^+] \) are \( E \)-nearly homotopic, and likewise the segments \( C_{ij}[y_{ij}^-, y_{ij}^+] \) and \( C_{i'j'}[y_{i'j'}^-, y_{i'j'}^+] \) are \( E \)-nearly homotopic, for any \( i, j, i', j' \in \{0, 1\} \);
(iii) the segments \( C_{0j}[x_{0j}^-, y_{0j}^+] \) and \( C_{1j}[x_{1j}^-, y_{1j}^+] \) are \( E \)-nearly homotopic;
(iv) the segments \( C_{i0}[y_{i0}^-, x_{i0}^+ + 1(C_{i0})] \) and \( C_{i1}[y_{i1}^-, x_{i1}^+ + 1(C_{i1})] \) are \( E \)-nearly homotopic;
(v) \( y_{00}^+ - x_{00}^- \geq R + K \) and \( x_{00}^+ + 1(C_{00}) - y_{00}^- \geq R + K \).
Then for $R > R_0$, we have
\[
\sum_{i,j=0,1} (-1)^{i+j} C_{ij} = 0
\]
in $\Pi_{10\varepsilon,R}$ homology.

Remark. Hypothesis (v) is satisfied provided that $y_{00}^+ - y_{00}^- \geq R$.

Proof. Set $K = 1 + E - \log \varepsilon$. For simplicity we write $I(C_{ij}) = I_{ij}$. We claim that we can find $x_{00} \in [x_{00}^- + \frac{K}{2}, x_{00}^+ - \frac{K}{2} + 1]$ and $y_{00} \in [y_{00}^- + \frac{K}{2}, y_{00}^+ - \frac{K}{2}]$ such that $y_{00} - x_{00} = R$. If $y_{00}^+ \leq x_{00}^- + R$, we let $x_{00} = x_{00}^- + \frac{K}{2}$ and $y_{00} = x_{00} + R$.

By the Convergence Lemma (Lemma 5.1), and by the choice of the constant $K$, we can find $x_{ij}^+ < x_{ij} < x_{ij}^-$ and $y_{ij}^- < y_{ij} < y_{ij}^+$ so that
\[
\text{dis}(C'_{ij}(x_{ij}^-), C'_{00}(x_{00}^-)), \text{dis}(C'_{ij}(y_{ij}^-), C'_{00}(y_{00}^-)) \leq \varepsilon,
\]
and the pairs of geodesic segments $C_{01}[x_{01}, y_{01}]$ and $C_{11}[x_{11}, y_{11}]$, and $C_{00}[y_{00}, x_{00} + 1]$ and $C_{11}[x_{11}, x_{11} + 1]$ are $\varepsilon$-nearly homotopic.

Let $I_{ij} = y_{ij} - x_{ij}$ and $J_{ij} = x_{ij} + 1 - y_{ij}$, so $I_{ij} + J_{ij} = I_{ij}$. Then $I_{00} = R$ and $J_{00} = I_{01} - R$, so $|J_{00} - R| < 2\varepsilon$.

Also, by the triangle inequality, we have $|I_{01} - R| = |I_{01} - I_{00}| < 2\varepsilon$. So
\[
|J_{01} - R| \leq |I_{01} - R| + |I_{01} - 2R| \leq 4\varepsilon.
\]
Then
\[
|J_{1j} - R| \leq |J_{0j} - R| + |J_{1j} - J_{0j}| < 6\varepsilon,
\]
so
\[
|I_{1j} - R| \leq |J_{1j} - R| + |I_{1j} - 2R| < 8\varepsilon.
\]
Therefore, we get $|I_{1j} - R| < 8\varepsilon$ for $i, j \in \{0, 1\}$.

We take
\[
\alpha_{00} \in \text{Conn}_{\varepsilon,R+\log 4}(\sqrt{-1}C'(x_{00}^-), -\sqrt{-1}C'(y_{00}^-)),
\]
and we let $\alpha_{ij}$ be the geodesic arc connecting $x_{ij}$ and $y_{ij}$ that is $\varepsilon$-nearly homotopic to $\alpha_{00}$ (see Figure 4). Then $\text{dis}(i(\alpha_{ij}), i(\alpha_{00})), \text{dis}(t(\alpha_{ij}), t(\alpha_{00})) \leq 2\varepsilon$. Therefore, because $\text{dis}(C'_{ij}(x_{ij}^-), C'_{00}(x_{00}^-)) \leq \varepsilon$ and $\text{dis}(C'_{ij}(y_{ij}^-), C'_{00}(y_{00}^-)) \leq \varepsilon$, we have
\[
\alpha_{ij} \in \text{Conn}_{3\varepsilon,R+\log 4}(\sqrt{-1}C'(x_{ij}^-), -\sqrt{-1}C'(y_{ij}^-)).
\]
Define $\Pi_{ij}$ as the pants generated from $C_{ij}$ by adding the third connection $\alpha_{ij}$. Denote by $A_{ij}$ and $B_{ij}$ the other two cuffs of $\Pi_{ij}$, oriented such that
\[
\partial \Pi_{ij} = C_{ij} - A_{ij} - B_{ij},
\]
where $A_{ij}$ is freely homotopic to the closed broken geodesic $C_{ij}[x_{ij}, y_{ij}]\alpha_{ij}^{-1}$, and $B_{ij}$ to $C_{ij}[y_{ij}, x_{ij} + 1]_0\alpha_{ij}$.
Applying Lemma 4.8, we obtain
\[ |l(A_{ij}) - 2R| < |I_{ij} - R| + 10 \varepsilon < 20 \varepsilon \]
and, similarly, \[ |l(B_{ij}) - 2R| < 20 \varepsilon \], so \( \Pi_{ij} \in \Gamma_{10\varepsilon,R} \). Finally, \( A_{i0} = A_{i1} \), and \( B_{0j} = B_{1j} \), so
\[ 0 = \sum_{i,j=0,1} (-1)^{i+j} \partial \Pi_{ij} = \sum_{i,j=0,1} (-1)^{i+j} \text{C}_{ij} t \]
in \( \Pi_{10\varepsilon,R} \) homology, which proves the lemma. \qed 

5.3. The Geometric Square Lemma.

Lemma 5.4 (Geometric Square Lemma). Let \( E, \varepsilon > 0 \). There exist constants \( K_1 = K_1(\mathcal{S},\varepsilon,E) > 0 \) and \( R_0(\mathcal{S},\varepsilon,E) > 0 \) with the following properties. Suppose that we are given four oriented geodesics \( C_{ij} \in \Gamma_{\varepsilon,R}, i,j = 0,1 \), and for each \( ij \) we are given 4 points \( x_{ij}^- < x_{ij}^+ < y_{ij}^- < y_{ij}^+ \). Assume that

(i) The inequalities \( x_{ij}^+ - x_{ij}^- > K_1 \), and \( y_{ij}^+ - y_{ij}^- > K_1 \), hold.

(ii) The segments \( C_{ij}[x_{ij}^-, x_{ij}^+] \) and \( C_{ij'}[x_{ij'}^-, x_{ij'}^+] \), are \( E \)-nearly homotopic, and likewise the segments \( C_{ij}[y_{ij}^-, y_{ij}^+] \) and \( C_{ij'}[y_{ij'}^-, y_{ij'}^+] \), are \( E \)-nearly homotopic, for any \( i,j,i',j' \in \{0,1\} \).

(iii) The segments \( C_{0j}[x_{0j}^-, y_{0j}^+] \) and \( C_{1j}[x_{1j}^-, y_{1j}^+] \) are \( E \)-nearly homotopic.

(iv) The geodesic segments \( C_{0j}[y_{0j}^-, x_{0j}^+ + l(C_{ij})] \) and \( C_{1j}[y_{1j}^-, x_{1j}^+ + l(C_{ij})] \) are \( E \)-nearly homotopic.

Then for \( R > R_0 \), we have
\[ \sum_{i,j=0,1} (-1)^{i+j} \text{C}_{ij} = 0, \]
in \( \Pi_{100\varepsilon,R} \) homology.
Proof. Below we use $L_0 = L_0(S, \varepsilon, E)$ and $K_0 = (S, \varepsilon, E)$ to denote two sufficiently large constants whose values will be determined in the course of the argument. The constant $Q_0$ can depend on $K_0$ and $L_0$. The constants $K_1$ and $R_0$ (from the statement of the GSL) can depend on $K_0$ and $L_0$ and $Q_0$. Each of these constants will be implicitly defined as a maximum of expressions in terms of constants that precede the given constant in the partial order of dependence that we just have described.

If we cannot apply the PGSL, then possibly interchanging the roles of the $x$’s and the $y$’s, we find that
\[
\begin{align*}
x_{00}^+ & \leq y_{00} - I_{00} + R + K(\varepsilon, E) \\
& \leq y_{00} - R + K(\varepsilon, E) + 1,
\end{align*}
\]
where $K = K(\varepsilon, E)$ is the constant from the previous lemma. We then let $y_{00} = y_{00} + Q_0$ and let $w_{00} = y_{00} - R$. (We assume that $Q_0 > K$.) Then

\[
w_{00} > x_{00}^+ + 10,
\]
provided $Q_0 > K + 11$, and

\[
y_{00} + Q_0 \leq y_{00} \leq y_{00}^+ + Q_0 - K_1,
\]
which implies

\[
y_{00}^+ + 2(E - \log \varepsilon) + 10 \leq y_{00} \leq y_{00}^+ - 2(E + \log \varepsilon) - 10,
\]
provided $Q_0 \geq 2(E - \log \varepsilon) + 10$ and $K_1 \geq Q_0 + 2(E - \log \varepsilon) + 10$.

Therefore, by the Convergence Lemma we can find $y_{ij}$ in the interval $[y_{ij}, y_{ij}^+]$ such that $\text{dis}(C_{ij}(y_{ij}), C_{00}(y_{00})) \leq \varepsilon$. We then let $w_{ij} = y_{ij} - R$

\[
\begin{align*}
geq & x_{ij}^+ + 10 \\
& (\text{provided } Q_0 > E + K + 12) \\
& \geq x_{ij}^- + K_1 \\
& \geq x_{ij}^- + 2(E - \log \varepsilon) + 10 \\
& (\text{provided } K_1 > 2(E - \log \varepsilon) + 10).
\end{align*}
\]

It follows from Lemma 5.1 that $C_i0[w_{i0}, y_{i0}]$ and $C_i1[w_{i1}, y_{i1}]$ are $\varepsilon$ $C^1$ nearly homotopic. (Two segments are $C^1$ nearly homotopic if the two initial and the two terminal vectors are $\varepsilon$ close in the tangent bundle respectively.) The point is that these two segments are contained into much larger segments that are $E$-nearly homotopic.

Let $(q, v) \in T^1S$, and take $\beta_{i0} \in \text{Conn}_{\varepsilon,L_0}(v, -\sqrt{-1}C_{i0}'(w_{i0}))$ (where we assume that $L_0 > L_0(\varepsilon, S)$ and $L_0(\varepsilon, S)$ is the constant from the Connection Lemma (Lemma 3.1)). We take $\alpha_{00} \in \text{Conn}_{\varepsilon,R+\log 4 - L_0}(\sqrt{-1}C_{00}'(y_{00}), v)$ (see Figure 5). Then we find $\alpha_{ij} \in \text{Conn}_{3\varepsilon, R+\log 4 - L_0}(\sqrt{-1}C_{ij}'(y_{ij}), v)$ and $\beta_{ij} \in$
Conn_{3\varepsilon,L_0}(v,-\sqrt{-1}C_{ij}'(w_{ij})) such that \alpha_{00} and \alpha_{ij} are \varepsilon-nearly homotopic and
\beta_{0i} and \beta_{ij} are 2\varepsilon-nearly homotopic for every i,j = 0,1.

Let \Pi_{ij} be the pair of pants generated by the geodesic segment
C_{ij}[w_{ij},y_{ij}], the broken geodesic segment \beta_{ij}^{-1}\alpha_{ij}^{-1} and, last but not least, the geodesic
segment \left(C_{ij}[y_{ij},w_{ij}+(l_{ij})]\right)^{-1}. The reader can verify that it is a topological
pair of pants.

Let A_{ij} be the closed geodesic freely homotopic to \alpha_{ij}\beta_{ij}C_{ij}[w_{ij},y_{ij}] and
let B_{ij} be the one for \left(C_{ij}[y_{ij},w_{ij}+(l_{ij})]\right)−1. Then ∂Π_{ij} = C_{ij} − A_{ij} − B_{ij}.

Using the second inequality from the remark just after the Sum of In-
efficiencies Lemma (see Lemma 4.8), we find that
\|l(A_{ij})−2R\| \leq 13\varepsilon and
\|l(B_{ij})−2R\| \leq 15\varepsilon. Hence \Pi_{ij} \in \Pi_{10\varepsilon,R}.

Observe that A_{i0} = A_{i1}, so
\sum_{i,j=0,1}(−1)^{i+j}C_{ij} − \sum_{i,j=0,1}(−1)^{i+j}∂Π_{ij} = \sum_{i,j=0,1}(−1)^{i+j}B_{ij}.

Let the \alpha_{ij}, a_{ij}, b_{ij}, and \beta_{ij} be real numbers and B_{ij} : \mathbb{R} \to B_{ij} be a parametriza-
tion of the geodesic B_{ij} so that B_{ij}(a_{ij}^+), B_{ij}(a_{ij}^-), B_{ij}(b_{ij}^+), and B_{ij}(b_{ij}^-) are the
projections of points q, C_{ij}(y_{ij}^+), C_{ij}(y_{ij}^-), and C_{ij}(x_{ij}^+) respectively onto the
geodesic B_{ij}. The points q, C_{ij}(y_{ij}^+), C_{ij}(x_{ij}^-), C_{ij}(x_{ij}^+) belong to the bro-
ked geodesic C_{ij}[y_{ij},w_{ij}+(l_{ij})\beta_{ij}^{-1}\alpha_{ij}^{-1}] and we project them to B_{ij} by choosing
lifts of B_{ij} and C_{ij}[y_{ij},w_{ij}+(l_{ij})\beta_{ij}^{-1}\alpha_{ij}^{-1}] in \mathbb{H}^2 that have the same endpoints
and then use the standard projection onto the lift of B_{ij}. 

Figure 5. The Geometric Square Lemma
It follows from the Convergence Estimate that each of \( q_i, C_{ij}(y_{ij}^+), C_{ij}(x_{ij}^-) \) and \( C_{ij}(x_{ij}^+) \) are within distance 1 of the corresponding projections on \( B_{ij} \). Then

\[
b_{ij}^+ - b_{ij}^- \geq x_{ij}^+ - x_{ij}^- - 2 \\
\geq K_0
\]

(provided \( K_1 > K_0 + 2 \))

and

\[
a_{ij}^+ - a_{ij}^- \geq R - L_0 - 3 + K_1 - Q_0 - E - 1 \\
\geq R + K_0
\]

(provided \( K_1 > K_0 + Q_0 + L_0 + E + 4 \)).

Assuming that \( K_0 \geq K(10\varepsilon, E + 2) \) (where \( K \) is the constant from the PGSL), we find that the differences \( b_{ij}^+ - b_{ij}^- \) and \( a_{ij}^+ - a_{ij}^- \) satisfy the lower bound from the PGSL. (Observe that \( b_{ij}^+ \) and \( b_{ij}^- \) are \( E \)-close and similarly for the \( a \)'s.) Also, the \( B_{ij} \)'s are in \( \Gamma_{10\varepsilon,R} \). So we apply the PGSL to show that

\[
\sum_{i,j=0,1} (-1)^{i+j} B_{ij} = 0
\]
in \( \Pi_{100\varepsilon,R} \) homology.

Here we explain why the assumptions of the PGSL are satisfied. For each \( i \), piecewise geodesics \( ai0-C_{i0}[y_{i0}, x_{i0}^+] \) and \( a1i-C_{i1}[y_{i1}, x_{i1}^+] \) are \( E \)-nearly homotopic; it follows from Lemma 4.1 that \( B_{i0}[a_i0^-, a_i0^+] \) and \( B_{i1}[a_i1^-, a_i1^+] \) are \( (E + 4) \)-nearly homotopic.

Likewise,

\[
C_{0j}[x_{0j}, w_{0j}] \beta_{0j}^{-1} \alpha_{0j}^{-1} C_{0j}[y_{0j}, y_{0j}^-]
\]

and

\[
C_{1j}[x_{1j}, w_{1j}] \beta_{1j}^{-1} \alpha_{1j}^{-1} C_{1j}[y_{1j}, y_{1j}^+]
\]

are \( E \)-nearly homotopic (because the individual segments are), and hence \( B_{0j}[b_{0j}^-, a_{0j}^+] \) and \( B_{1j}[b_{1j}^-, a_{1j}^+] \) are.

\[\square\]

Randomization (Randomization remarks for the GSL). Let \( \varepsilon, E > 0 \). Every constant \( K \) below may depend only on \( \varepsilon \), \( S \) and \( E \).

Below we will define a partial map \( g : \left( \bar{\Gamma}_{1,R} \right)^4 \rightarrow \mathbb{R} \Pi_{100\varepsilon,R} \) such that

(i) \( g \) is defined on any input \( (C_{ij}, x_{ij}^+, y_{ij}^+) \) that satisfies hypotheses (i)–(iv) of GSL;

(ii) \( \sum(-1)^{i+j} C_{ij} = \partial g(C_{ij}, x_{ij}^+, y_{ij}^+) \);

(iii) \( g \) is \( K \)-semirandom with respect to measures classes \( \Sigma_{\Gamma} \) on \( \left( \bar{\Gamma}_{1,R} \right)^4 \) and \( \sigma_{\Pi} \) on \( \Pi_{1,R} \).
We first define a partial function $g_0 : (\tilde{\Gamma}_{1,R})^4 \to \mathbb{R} \Pi_{10\varepsilon,R}$ that is defined on inputs $(C_{ij}, x_{ij}^{\pm}, y_{ij}^{\pm})$ that satisfy the extra hypothesis (v) from the PGSL. Given such an input, we follow the construction of the PGSL to construct $C_{ij}$ from the old ones, the map $(C_{ij}, x_{ij}^{\pm}, y_{ij}^{\pm}) \to (C_{ij}, x_{ij}, y_{ij})$ is $K$-semirandom as a partial map from $(\tilde{\Gamma}_{1,R})^4$ to $(\tilde{\Gamma}_{1,R})^4$, with respect to the measure classes $\Sigma_{\Gamma}^{\varepsilon_4}$ on $(\tilde{\Gamma}_{1,R})^4$ and $\Sigma_{\Gamma}^{\varepsilon_4}$ on $(\tilde{\Gamma}_{1,R})^4$.

Then we take a random third connection

$$\alpha_{00} \in \text{Conn}_{\varepsilon,R+\log 4}(\sqrt{-IC_{00}'}(x_{00}), -\sqrt{-IC_{00}'}(y_{00})).$$

Likewise for $\alpha_{ij}$. Adding the third connection $\alpha_{ij}$ to $C_{ij}$ we obtain the pants $\Pi_{ij}(\alpha_{ij})$. We claim that distinct $\alpha_{ij}$ lead to distinct pants $\Pi_{ij}(\alpha_{ij})$. The third connection $\alpha_{ij}$ is $\varepsilon$-close to the unique simple geodesic arc on $\Pi_{ij}(\alpha_{ij})$ that is orthogonal to $\gamma_{ij}$ at both ends. On the other hand, no two distinct $\alpha_{ij}$ are $\varepsilon$-close, so assuming that the injectivity radius of the surface $S$ is at least $2\varepsilon$, we find that distinct $\alpha_{ij}$ give distinct $\Pi_{ij}(\alpha_{ij})$.

So, for each input $(C_{ij}, x_{ij}, y_{ij})$, by adding a random third connection $\alpha_{ij}$ we construct the pants $\Pi_{ij}(\alpha_{ij})$. So far, we have been using the term “random” to mean arbitrary. In these randomization remarks we will also interpret the phrase “a random element of a finite set $S$” as “the random element of $\mathbb{R} S$,” namely $\frac{1}{|S|} \sum_{x \in S} x$.

We can then think of every map $f : S \to T$ that we have implicitly constructed in the text as the associated linear map $f : \mathbb{R} S \to \mathbb{R} T$ defined by $f(\sum a_i x_i) = \sum a_i f(x_i)$. So, for example, we let

$$\overline{\alpha}_{ij} \in \text{Conn}_{\varepsilon,R+\log 4}(\sqrt{-IC_{ij}'}(x_{ij}), -\sqrt{-IC_{ij}'}(y_{ij}))$$

be the random element of

$$\text{Conn}_{\varepsilon,R+\log 4}(\sqrt{-IC_{ij}'}(x_{ij}), -\sqrt{-IC_{ij}'}(y_{ij})),$$

and then $\Pi_{ij}(\overline{\alpha}_{ij})$ is the image of $\overline{\alpha}_{ij}$ by the linear form of the map $\alpha_{ij} \to \Pi_{ij}(\alpha_{ij})$.

In this manner we have constructed a partial map from $\tilde{\Gamma}_{1,R} \to \Pi_{1,R}$ (defined by $(C_{ij}, x_{ij}, y_{ij}) \to \Pi_{ij}(\alpha_{ij})$; compare with Lemma 3.2), and we claim that it is $K$-semirandom with respect to $\Sigma_{\Pi}$ and $\sigma_{\Pi}$. To verify this claim we need to show that for any given pants $\Pi \in \Pi_{1,R}$, the weight of $\Pi$ is at most $K e^{-3R}$. Let $C$ be a cuff of $\Pi$, and choose points $x, y \in C$ that lie in certain unit length intervals on $C$. Let $\text{Conn}$ be the set of all good third connections between $x$ and $y$. (By this we mean all connections $\alpha$ so that $C$ and $\alpha$ produce a pair of pants in $\Pi_{1,R}$.) The set $\text{Conn}$ has approximately $e^{R-K}$ elements. Moreover, there is a unique third connection $\alpha \in \text{Conn}$ so that $\alpha$ and $C$ yield
the given pair of pants \( \Pi \). So, the total weight of \( \Pi \) is at most \( e^{K-R} \) times the total weight for the three choices of \( C \in \partial \Pi \) (with associated unit intervals), and we conclude that the total weight for \( \Pi \) is at most \( 3e^{K-R}e^{-2R} = Ke^{-3R} \).

Also, the map \((\Pi_{ij})_{i,j \in \{0,1\}} \mapsto \sum (-1)^{i+j} \Pi_{ij}\) is of course 4-semirandom from \((\Pi_{1,R}^1, \Sigma_{\Pi}^{\otimes 4})\) to \((\mathbb{R}\Pi_{1,R}, \Sigma_{\Pi})\). Composing the above maps we construct the map \(g_0\) and see that \(g_0\) is \(K\)-semirandom.

For the general case, similarly as above we first define the map

\[
h : (C_{ij}, x_{ij}^{\pm}, y_{ij}^{\pm}) \to \mathbb{R}\Pi_{1,R}
\]

according to our second construction, on every input \((C_{ij}, x_{ij}^{\pm}, y_{ij}^{\pm})\) that satisfies conditions (i)–(iv) of the GSL, but not condition (v) of the PGSL.

We construct \(y_{ij}\) and \(w_{ij}\) as before. The map \((C_{ij}, x_{ij}^{\pm}, y_{ij}^{\pm}) \to (C_{ij}, y_{ij}, w_{ij})\) is \(K\)-semirandom. Then we find the connections \(\alpha_{ij}\) and \(\beta_{ij}\). There are at least \(e^{R-K}\) of the \(\alpha_{ij}\) (we only fix a single \(\beta_{ij}\)), and each third connection \(\alpha_{ij}\beta_{ij}\) leads to a new pair of pants \(\Pi_{ij}(\alpha_{ij}\beta_{ij})\). Let \(N\) denote the number of connections \(\alpha_{ij}\). (By construction, the number \(N\) does not depend on \(i\) and \(j\).) This defines the map

\[
h(C_{ij}, x_{ij}^{\pm}, y_{ij}^{\pm}) = \sum \frac{1}{N} \Pi_{ij}(\alpha_{ij}\beta_{ij}),
\]

and we can verify that \(h\) is \(K\)-semirandom.

Then we observe that \(\partial_B : \Pi_{ij} \to B_{ij}\) formed by taking the appropriate boundary curve of the \(\Pi_{ij}\) we constructed is \(K\)-semirandom, so the induced map \(\bar{h} : (C_{ij}, x_{ij}^{\pm}, y_{ij}^{\pm}) \to (B_{ij}, a_{ij}, b_{ij})\) is as well. So the map \(g_1\) defined by

\[
g_1(C_{ij}, x_{ij}^{\pm}, y_{ij}^{\pm}) = \sum (-1)^{i+j} \Pi_{ij} + g_0(B_{ij}, a_{ij}, b_{ij})\]

is \(K\)-semirandom, and hence

\[
g = g_0 + g_1\]

is as well.

6. The Algebraic Square Lemma

We prove the Algebraic Square Lemma, which will be used in almost all of our subsequent identities in the Good Pants Homology. In particular, it will allow us to encode an element of \(\pi_1(S,\ast)\) as a sum of good pants and then prove that the encoding of products of elements of \(\pi_1(S,\ast)\) is the sum of their encodings.

61. Notation. Let \(\ast \in S\) denote a point that we fix once and for all.

By \(\pi_1(S,\ast)\) we denote the fundamental group of a pointed surface. If \(A \in \pi_1(S,\ast)\), we let \(\cdot A\) be the geodesic segment from \(\ast\) to \(\ast\) homotopic to \(A\). By \([A]\) we denote the closed geodesic on \(S\) that is freely homotopic to \(\cdot A\). If \(A_1, \ldots, A_n \in \pi_1(S,\ast)\) we let \(\cdot A_1 \cdot A_2 \cdots \cdot A_n\) be the piecewise geodesic arc that is the concatenation of the arcs \(\cdot A_i\). We let \([\cdot A_1 \cdot A_2 \cdots \cdot A_n]\) be the closed piecewise geodesic that arises from the arc \(\cdot A_1 \cdot A_2 \cdots \cdot A_n\) by noticing that the starting and the ending point of \(\cdot A_1 \cdot A_2 \cdots \cdot A_n\) are the same. By
I([A]) is the length of the closed geodesic [A]. By I(·A·) we mean of course the length of the geodesic arc ·A·, and in general by I(·A₁ · · · Aₙ·) the length of the corresponding piecewise geodesic arc.

Remark. Observe that for any \( X_i \in \pi_1(S, \ast) \), \( i = 0, \ldots, n - 1 \), the closed geodesics \([X_j X_{j+1} \ldots X_{n+j-1}]\) are one and the same. (We are taking the indices modulo \( n \).) We will call this rotation and often use it without warning.

We remind the reader that \( \cdot AB \cdot \) is a geodesic arc from \( \ast \) to \( \ast \) representing \( AB \), while \( \cdot A \cdot B \cdot \) is a concatenation of two geodesic arcs. Similarly, \( \cdot AB \cdot C \cdot \) is a concatenation of two geodesic arcs, while \( \cdot A \cdot B \cdot C \cdot \) is a concatenation of three, and so on.

In particular, we have the following statements about the inefficiency function:

\[
I(\cdot A_1 \cdot \ldots \cdot A_n \cdot) = \sum l(\cdot A_i \cdot) - l(\cdot A_1 A_2 \ldots A_n \cdot)
\]

and

\[
I([\cdot A_1 \cdot \ldots \cdot A_n \cdot]) = \sum l(\cdot A_i \cdot) - l([A_1 \ldots A_n]).
\]

Notice that we may have (and will usually have)

\[
I([\cdot A_1 \cdot \ldots \cdot A_n \cdot]) > I(\cdot A_1 \cdot \ldots \cdot A_n \cdot).
\]

6.2. The Algebraic Square Lemma (the ASL). The following is the Algebraic Square Lemma.

Lemma 6.1 (Algebraic Square Lemma). Let \( \varepsilon, \Delta > 0 \). There exist constants \( K(S, \varepsilon, \Delta) = K \) and \( R_0 = R_0(S, \varepsilon, \Delta) \) so that for \( R > R_0 \), the following holds. Let \( A_i, B_i, U, V \in \pi_1(S, \ast), i = 0, 1, \) be such that

(i) \(|I(\cdot A_i UB_j V \cdot) - 2R| < 2 \varepsilon, i, j = 0, 1; \)
(ii) \(|I([\cdot A_i \cdot U \cdot B_j \cdot V \cdot])| < \Delta; \)
(iii) \(|I(\cdot U \cdot), l(\cdot V \cdot) > K.\)

Then

\[
\sum_{ij} (-1)^{i+j}[A_i UB_j V] = 0
\]

in \( \Pi_{100 \varepsilon, R} \) homology.

Proof. For each \( i, j \in \{0, 1\} \), we project the closed piecewise geodesic \([\cdot A_i \cdot U \cdot B_j \cdot V \cdot]\) onto the closed geodesic \( \gamma_{ij} = [A_i UB_j V] \). By Lemma 4.6 we find that each appearance of \( \ast \) is moved at most distance \( E = \frac{\Delta}{2} + 1 \) by the projections. Let \( \gamma_{ij}(x_{ij}^\pm) \) and \( \gamma_{ij}(y_{ij}^\pm) \) be the projections of \( \ast \) on \( \gamma_{ij} \) before and after \( U \), and before and after \( V \), respectively. Then providing that our \( K \) is at least \( 2E \) plus the corresponding constant from the GSL, we have \( x_{ij} < x_{ij}^+ < y_{ij} < y_{ij}^+ < x_{ij}^- + l(\gamma_{ij}) \) and the hypotheses of the Geometric Square
Lemma. We conclude that 

$$\sum_{ij} (-1)^{i+j} [A_i U B_j V] = 0$$

in $\Pi_{100\varepsilon,R}$ homology. \(\square\)

Randomization (Randomization remarks for the ASL). Let $\varepsilon, \delta > 0$. By $K$ we denote any constant that may depend only on $\varepsilon$, $S$, and $\Delta$.

Below we will define a partial map $f : G^2 \times G \times G^2 \times G \to \mathbb{R} \Pi_{1, R}$ such that

(i) $f$ is defined on any input $(A_i, U, B_j, V)$ that satisfies the assumptions of the ASL;
(ii) $\sum (-1)^{i+j} [A_i U B_j V] = \partial f(A_i, U, B_j, V);$ 
(iii) $f$ is $K$-semirandom with respect to the classes of measures $\Sigma_G^{22} \times \Sigma_G \times \Sigma_G^{22} \times \Sigma_G$ on $G^2 \times G \times G^2 \times G$ and $\sigma_\Pi$ on $\Pi_{1, R}$.

Let $h$ be a partial map $h : G^2 \times G \times G^2 \times G \to \left(\tilde{\Gamma}_{1, R}\right)^4$ defined by letting $h(A_i, U, B_j, V) = (C_{ij}, x_{ij}^+, y_{ij}^+)$, where $C_{ij} = [A_i U B_j V]$, and $x_{ij}^+$ and $y_{ij}^+$ are the points on the parametrizing torus for $C_{ij}$ such that the points $C_{ij}(x_{ij}^+)$ and $C_{ij}(y_{ij}^+)$ are the corresponding projections of the four copies of the base point $*$ (which belong to the closed piecewise geodesic $[\cdot A_i \cdot U \cdot B_j \cdot V \cdot]$) to the closed geodesic $C_{ij}$. (These projections were defined above.) It follows from Lemmas 10.2 and 10.3 that $h$ is $K$-semirandom. Let $g : \left(\tilde{\Gamma}_{1, R}\right)^4 \to \mathbb{R} \Pi_{1, R}$ be the $K$-semirandom map from the previous section. (See the randomization remarks for the GSL.) Then $f = g \circ h$ is $K$-semirandom.

6.3. The Sum of Inefficiencies Lemma in the algebraic notation. The following lemma follows from Lemma 4.8.

**Lemma 6.2** (Sum of Inefficiencies Lemma in the algebraic notation). Let $\varepsilon, \Delta > 0$ and $n \in \mathbb{N}$. There exists $L = L(\varepsilon, \Delta, n) > 0$ such that if $U_1, \ldots, U_{n+1} = U_1, X_1, \ldots, X_n \in \pi_1(S, *)$, and $I(\cdot U_i \cdot X_i \cdot U_{i+1}) \leq \Delta$, and $I(\cdot \text{ot} U_i \cdot) \geq L$, then

$$\left| I(\cdot U_1 \cdot U_2 \cdot X_1 \cdot U_2 \cdot X_2 \cdot \ldots \cdot U_n \cdot X_n \cdot) - \sum_{i=1}^n I(\cdot U_i \cdot X_i \cdot U_{i+1} \cdot) \right| \leq \varepsilon.$$ 

Remark. In particular, we can leave out the $X$’s in the above lemma and write

$$\left| I(\cdot U_1 \cdot U_2 \cdot \ldots \cdot U_n \cdot) - \sum_{i=1}^n I(\cdot U_i \cdot U_{i+1} \cdot) \right| \leq \varepsilon,$$
where \( \theta \).

Likewise, such that if \( U \) and \( \epsilon \) large enough, we have

\[
\left| I([ urm U_1 \cdot U_2 \cdot \ldots \cdot U_n \cdot \cdot]) - \sum_{i=1}^{n} I(\theta_i) \right| \leq 2\epsilon,
\]

where \( \theta_i = \Theta(t(U_i \cdot \cdot), i(U_{i+1} \cdot \cdot)). \)

Similarly, the following lemma follows from Lemma 4.9.

**Lemma 6.3.** Let \( \epsilon, \Delta > 0 \) and \( n \in \mathbb{N} \). There exists \( L = L(\epsilon, \Delta, n) > 0 \) such that if \( U_1, \ldots, U_{n+1} = U \in \pi_1(S, \cdot) \) and \( X_1, \ldots, X_{n_1}, \ldots, X_{n_1}, \ldots, X_{n_{n_j}} \in \pi_1(S, \cdot) \), and \( I(U_i \cdot X_i \cdot U_{i+1}) \leq \Delta \), and \( I(U_i \cdot \cdot) \geq L \), then

\[
\left| I([ urm U_1 \cdot X_{n_1} \cdot \ldots \cdot X_{n_{n_j}} \cdot \cdot U \cdot X_{n_1} \cdot \ldots \cdot X_{n_{n_j}} \cdot \cdot]) - \sum_{i=1}^{n} I(U_i \cdot X_i \cdot \cdot U_{i+1} \cdot \cdot) \right| \leq \epsilon.
\]

Finally, we have the Flipping Lemma.

**6.4. The Flipping Lemma.** For \( X \in \pi_1(S, \cdot) \), we let \( \bar{X} = X^{-1} \) denote the inverse of \( X \).

**Lemma 6.4 (Flipping Lemma).** Let \( \epsilon, \Delta > 0 \). There exists a constant \( L = L(\epsilon, \Delta) > 0 \) with the following properties. Suppose \( A, B, T \in \pi_1(S, \cdot) \), and

\[
I(\cdot T \cdot A \cdot \bar{T} \cdot \cdot), I(\cdot \bar{T} \cdot B \cdot T \cdot \cdot) \leq \Delta,
\]

and \( I(\cdot T \cdot \cdot) \geq L \). Then

\[
|I([ urm T \cdot A \cdot \bar{T} \cdot B \cdot \cdot]) - I([ urm T \cdot A \cdot \bar{T} \cdot B \cdot T])| < \epsilon,
\]

and therefore

\[
|I([TA\bar{T}B]) - I([TA\bar{T}B])| < \epsilon.
\]

**Proof.** By the Long Segment Lemmas,

\[
|I(\cdot T \cdot A \cdot \bar{T} \cdot B \cdot T \cdot \cdot) - I(\cdot T \cdot A \cdot \bar{T} \cdot \cdot) - I(\cdot \bar{T} \cdot B \cdot T \cdot \cdot)| < \frac{\epsilon}{4},
\]

and

\[
|I(\cdot T \cdot A \cdot \bar{T} \cdot B \cdot \cdot) - I(\cdot T \cdot A \cdot \bar{T} \cdot B \cdot T \cdot \cdot)| < \frac{\epsilon}{4}.
\]

Likewise,

\[
|I(\cdot T \cdot A \cdot \bar{T} \cdot B \cdot T \cdot \cdot) - I(\cdot T \cdot A \cdot \bar{T} \cdot \cdot) - I(\cdot \bar{T} \cdot B \cdot T \cdot \cdot)| < \frac{\epsilon}{4},
\]

and

\[
|I(\cdot T \cdot A \cdot \bar{T} \cdot B \cdot \cdot) - I(\cdot T \cdot A \cdot \bar{T} \cdot B \cdot T \cdot \cdot)| < \frac{\epsilon}{4}.
\]
But \( I([\cdot \cdot T \cdot A \cdot \bar{T} \cdot]) = I([\cdot \cdot T \cdot \bar{A} \cdot \bar{T} \cdot]) \), because \( \cdot \cdot T \cdot A \cdot \bar{T} \cdot \) is the same as \( \cdot \cdot T \cdot \bar{A} \cdot \bar{T} \cdot \) with reversed orientation. So

\[
|I([\cdot \cdot T \cdot A \cdot \bar{T} \cdot B \cdot]) - I([\cdot \cdot T \cdot \bar{A} \cdot \bar{T} \cdot B \cdot])| < \varepsilon.
\]

Similarly, we conclude \(|l([\cdot \cdot T \cdot A \cdot \bar{T} \cdot B \cdot]) - l([\cdot \cdot T \cdot \bar{A} \cdot \bar{T} \cdot B \cdot])| < \varepsilon.\)

\[\square\]

7. Applications of the Algebraic Square Lemma

In this section we will describe the encoding of an element \( \bar{A} \) of \( \pi_1(\ast) \) as a sum \( A_T \) of good pants. (The encoding depends on a choice of a sufficiently large element \( T \) of \( \pi_1(\ast) \).)

In brief, we let

\[
A_T = \frac{1}{2} (\lfloor (A \bar{T} B) \rfloor - \lfloor (\bar{A} \bar{T} B) \rfloor)
\]

for suitable \( B \), and we then observe that the Algebraic Square Lemma implies that different choices of \( B \) give the same element of the good pants homology.

We can then easily prove that

\[
\lfloor A T B \rfloor = A_T + B_T,
\]

which we call the Two-Part Itemization Lemma. We want to go one step further and prove that

\[
\lfloor A T B C D \rfloor = A_T + B_T + C_T + D_T
\]

for suitable \( A, B, C, D \) and \( T \).

It turns out that in order to prove this Four-Part Itemization Lemma, we must first prove that

\[
\lfloor A T B C D \rfloor = \lfloor A T D C T \rfloor.
\]

This is indeed the most difficult lemma of this section.

We would then be able to go ahead and prove a Six-Part Itemization Lemma and so forth, but the Four-Part Itemization Lemma is sufficient for our purposes.

We state several results and definitions (notably the definition of \( A_T \) in the next lemma) that depend on an element \( T \in \pi_1(\ast) \) and \( \Delta > 0 \). We treat both \( T \) and \( \Delta \) as parameters, and the exact value of both \( T \) and \( \Delta \) (which are then used in the proof of the main theorem) will be determined in Section 9.

7.1. The definition of \( A_T \). For \( A, T \in \pi_1(\ast) \) and \( \varepsilon, R > 0 \), we let \( \mathcal{F}_{\Gamma_{\varepsilon, R}}(A, T) \) be the set of all \( B \in \pi_1(\ast) \) such that \( \lfloor A T B \rfloor, \lfloor A T B \rfloor \in \Gamma_{\varepsilon, R} \) and \( I(\cdot \cdot T \cdot B \cdot T \cdot) < 1 \).

**Lemma 7.1.** Let \( \varepsilon, \Delta > 0 \). There exists a constant \( L = L(\ast, \varepsilon, \Delta) \) such that if \( A, T \in \pi_1(\ast) \) and \( I(\cdot \cdot T \cdot A \cdot \bar{T} \cdot) \leq \Delta \), and \( I(\cdot \cdot \bar{A} \cdot) \geq L \), and \( 2R - I(\cdot \cdot A \cdot) - 2I(\cdot \cdot T \cdot) \geq L \), then
(i) $\mathcal{F}_{\text{Conn},R}^\varepsilon(A,T)$ is nonempty, and $\log |\mathcal{F}_{\text{Conn},R}^\varepsilon(A,T)| \geq 2R - 1(\cdot A \cdot) - 2I(\cdot T \cdot) - \Delta - L$;
(ii) $[TA\bar{T}B] - [T\bar{A}\bar{T}B] = [TA\bar{T}B'] - [T\bar{A}\bar{T}B']$ in $\mathbf{H}_{100\varepsilon,R}$ homology for any $B, B' \in \mathcal{F}_{\text{Conn},R}^\varepsilon(A,T)$.

We then let

$$A_T = \frac{1}{2} ([TA\bar{T}B] - [T\bar{A}\bar{T}B])$$

for a random $B \in \mathcal{F}_{\text{Conn},R}^\varepsilon(A,T)$.

**Remark.** Part (ii) of Lemma 7.1 implies that for any $B \in \mathcal{F}_{\text{Conn},R}^\varepsilon(A,T)$, we have

$$A_T = \frac{1}{2} ([TA\bar{T}B] - [T\bar{A}\bar{T}B])$$

in $\mathbf{H}_{100\varepsilon,R}$ homology. Also, it is important to note that $[A]$ is equal to $A_T$ in the standard homology $\mathbf{H}_1$.

**Proof.** Suppose that $\cdot B \cdot \in \text{Conn}_{\varepsilon,R'}(-i(\cdot T \cdot), i(\cdot T \cdot))$, where $R' = 2R - I(\cdot A \cdot) - 2I(\cdot T \cdot) - I(\cdot T \cdot A \cdot T \cdot)$. The set $\text{Conn}_{\varepsilon,R'}(-i(\cdot T \cdot), i(\cdot T \cdot))$ will be nonempty (by the Connection lemma (Lemma 3.1)) provided $L$ is large. Then, by the Sum of Inefficiencies Lemma,

$$|I([T\bar{A}\bar{T}B]) - 2R| < \varepsilon + O(\varepsilon^2)$$

and

$$|I([TA\bar{T}B]) - 2R| < \varepsilon + O(\varepsilon^2),$$

provided $I(\cdot T \cdot)$ is large. Thus, with slight abuse of notation, we have

$$\text{Conn}_{\varepsilon,R'}(-i(\cdot T \cdot), i(\cdot T \cdot)) \subset \mathcal{F}_{\text{Conn},R}^\varepsilon(A,T),$$

and

$$\log |\text{Conn}_{\varepsilon,R'}(-i(\cdot T \cdot), i(\cdot T \cdot))| \geq 2R - 1(\cdot A \cdot) - 2I(\cdot T \cdot) - L$$

if $L$ is large, so we have proved statement (i) of the lemma.

Again, by the Sum of Inefficiencies Lemma, the inefficiency of the piecewise geodesic $[\cdot T \cdot A \cdot T \cdot B \cdot ]$ is at most $\Delta + 2$. Then statement (ii) follows, provided $L$ (and hence $I(\cdot T \cdot)$) is large, from the Algebraic Square Lemma.

**Randomization** (Randomization remarks for $A_T$). All constants $K$ may depend only on $\varepsilon$, $\Delta$, and $S$ and $T \in \pi_1(S,*)$.

Letting $B_A \in \mathbb{R}G$ denote the random element of $\mathcal{F}_{\text{Conn},R}^\varepsilon(A,T)$, we consider the map $A \rightarrow B_A$ from $G$ to $\mathbb{R}G$. If $I(\cdot A \cdot) \in [a, a+1]$, we find that $I(\cdot B \cdot) \in [L_a, R_a]$, where $L_a = 2R - a - 2I(\cdot T \cdot) - \Delta - 4$ and $R_a = 2R - a - 2I(\cdot T \cdot)$ for all $B \in \mathcal{F}_{\text{Conn},R}^\varepsilon(A,T)$. Because

$$|\mathcal{F}_{\text{Conn},R}^\varepsilon(A,T)| > e^{L_a - L}$$
(where \( L = L(\varepsilon, S) \) from the Connection Lemma), and \( \sigma_a(G) \leq K \) (see the appendix for the definition of \( \sigma_a \)), we find that for any \( X \in G \),

\[
(A \to B_A)_*\sigma_a(X) \leq Ke^{L-L_a}
\]

if \( I(\cdot, X \cdot) \in [L_a, R_a], \) and \( (A \to B_A)_*\sigma_a(X) = 0 \) otherwise. This implies

\[
(A \to B_A)_*\sigma_a \leq K \sum_{k=\lfloor L_a \rfloor}^{R_a} e^{k+L-L_a} \sigma_k,
\]

which in turn implies that the map \( A \to B_A \) is \( K \)-semirandom with respect to \( \Sigma_G \) and \( \Sigma_G^3 \).

We define \([A\bar{T}B_A T]\) by

\[
[A\bar{T}B_A T] = \frac{1}{|FConn_{\varepsilon,R}(A,T)|} \sum_{B \in FConn_{\varepsilon,R}(A,T)} [A\bar{T}B_T].
\]

The map \( A \to (A,T) \) is \( e^{1(T\cdot)} \)-semirandom by the remark stated just before the Principles of randomization section in the appendix.

Then the partial maps from \( G \) to \( \mathbb{R}G^4 \) defined by \( A \to (A, \bar{T}, B_A, T) \) and \( A \to (A, \bar{T}, B_A, T) \) are \( Ke^{21(T\cdot)} \)-semirandom with respect to \( \Sigma_G \) and \( \Sigma_G^3 \), and hence the map

\[
A \to A_T = \frac{1}{2} ([A\bar{T}B_A T] - [A\bar{T}B_A T])
\]

is \( Ke^{21(T\cdot)} \)-semirandom with respect to \( \Sigma_G \) and \( \sigma_T \).

The map \( (A, B') \to (A, B_A, B') \) is \( K \)-semirandom with respect to \( \Sigma_G^3 \) and \( \Sigma_G^3 \), and \( (A, B_A, B') \to (A, \bar{T}, B_A, B', T) \) is \( Ke^{21(T\cdot)} \)-semirandom with respect to \( \Sigma_G^3 \) and \( \Sigma_G^3 \times \Sigma_G \times \Sigma_G^3 \times \Sigma_G \).

Also, by the Algebraic Square Lemma, the map \( (A, \bar{T}, T, B, B', T) \to \Pi \in \mathbb{R}\Pi_{1,R} \), such that

\[
\partial \Pi = [T\bar{A}T\bar{B}] - [T\bar{A}T\bar{B}] - [T\bar{A}T\bar{B}] + [T\bar{A}T\bar{B}],
\]

is \( K \)-semirandom from \( G^2 \times G \times G^2 \times G \) to \( \Pi_{1,R} \), with respect to the measure classes \( \Sigma_G^2 \times \Sigma_G \times \Sigma_G^2 \) and \( \Sigma_G^3 \) and \( \sigma_{\Pi} \). Composing the above mappings we find a \( Ke^{21(T\cdot)} \)-semirandom map \( g : G^2 \to \mathbb{R}\Pi_{1,R} \) such that

\[
\partial g(A, B') = A_T - \frac{1}{2} ([T\bar{A}T\bar{B}] - [T\bar{A}T\bar{B}]).
\]

Remark. At the end of the paper we will see that \( T \) and \( \Delta \) only depend on \( S \) and \( \varepsilon \).

7.2. The Two-part Itemization Lemma. The following lemma is a corollary of the previous one, and we refer to it as the Two-part Itemization Lemma.

Lemma 7.2 (Two-part Itemization Lemma). Let \( \varepsilon, \Delta > 0 \). There exists a constant \( L = L(S, \varepsilon, \Delta) > 0 \) such that for any \( A, B, T \in \pi_1(S, \ast) \) such that
\[ [TA\bar{T}B] \in \Gamma_{\epsilon, R}, \text{ we have } [TA\bar{T}B] = A_T + B_T \text{ in } \Pi_{200\epsilon, R} \text{ homology, provided that } I(\cdot T \cdot), I(\cdot A \cdot), I(\cdot B \cdot) > L \text{ and } I([\cdot T \cdot A \cdot T \cdot B \cdot]) \leq \Delta. \]

Proof. It follows from Lemma 6.4 that \([TA\bar{T}B] \in \Gamma_{2\epsilon, R}. \) In order to apply Lemma 7.1, we need an upper bound on \(I(\cdot T \cdot A \cdot T \cdot)\) and a lower bound on \(2R - I(\cdot A \cdot - 2I(\cdot T \cdot). \) These follow from \(I(\cdot B \cdot), I(\cdot T \cdot) \geq L, I([\cdot T \cdot A \cdot T \cdot B \cdot]) \leq \Delta, \) and the Sum of Inefficiencies Lemma.

We observe

\[
[TATB] = \frac{1}{2}([TATB] - [BTA\bar{T}]) \\
= \frac{1}{2}([TATB] - [T\bar{A}TB]) \\
= \frac{1}{2}([TATB] - [T\bar{A}B]) \\
+ \frac{1}{2}([\bar{T}BT\bar{A}] - [\bar{T}BTA]) \\
= A_T + B_T
\]

in \(\Pi_{200\epsilon, R} \) homology. \(\square\)

Randomization (Randomization remarks for the Two-part Itemization Lemma). We have implicitly defined a map \(g : G^2 \to \mathbb{R}\Pi_{100\epsilon, R}\) such that \(\partial g(A, B) = A_T + B_T - [ATBT]. \) The map \(g\) is \(K\epsilon^{2|\cdot T \cdot|}. \) semirandom with respect to \(\Sigma_G \times \Sigma_G\) and \(\sigma_{\Pi}\).

Remark. In fact it should be true that

\[ [TA_1\bar{B}_1 \ldots TA_n\bar{B}_n] = \sum_{i=1}^{n}(A_i)_T + (B_i)_T, \tag{10} \]

provided \(I(\cdot T \cdot)\) is large given \(I(TA_i\bar{T})\) and \(I(\bar{T}B_iT). \) Above we proved this when \(n = 1\) (provided \(I(\cdot A \cdot)\) and \(I(\cdot B \cdot)\) are large), and we will prove it in the rest of this section for \(n = 2, \) using the ADCB lemma, which we prove next. The general case can be proved by induction using the cases \(n = 1\) and \(n = 2\) (but we will only need this statement for \(n = 1, 2\)).

Remark. We also observe that under the usual conditions we have \(A_{TU} = A_U\) in \(\Pi_{100\epsilon, R} \) homology. This follows from the fact that \(2A_{TU} = [TU\bar{A}\bar{T}B] + [TUA\bar{T}B] = [UA\bar{T}BT] + [UA\bar{T}\bar{B}T] = 2A_U.\)

7.3. The ADCB Lemma.

Claim. Let \(\delta, \Delta > 0. \) There exists \(L = L(\Delta, \delta) > 0\) with the following properties. Let \(A_i, B_i, T \in \pi_1(S, \ast), \) \(i = 0, 1. \) If \(I(\cdot T \cdot) > L\) and \(I(\cdot \bar{T} \cdot A_i \cdot T \cdot),\)
Lemma. Since by rotation \( V \) and \( \Gamma \) in \( \Pi \) the curves in question are in \( \Gamma \), then \( \|A_i \cdot T \cdot \overline{A_i} \cdot T\| < \delta \).

Proof. By the Sum of Inefficiencies Lemma we have that

\[
I([A_0 \cdot B_0 \cdot \overline{T} \cdot A_1 \cdot T \cdot B_1 \cdot \overline{T}])
\]

is close to

\[
\sum_{i=0,1} I(\cdot \overline{T} \cdot A_i \cdot T \cdot \cdot) + I(\cdot T \cdot B_j \cdot \overline{T} \cdot).
\]

By the definition of inefficiency, the number

\[
\sum I(\cdot \overline{T} \cdot A_i \cdot T \cdot \cdot) + I(\cdot T \cdot B_j \cdot \overline{T} \cdot) - 4I(\cdot T \cdot \cdot) - 4I(\cdot A_i) - 4I(B_i)
\]

is small in absolute value. The claim now follows from the Sum of Inefficiencies Lemma. \( \square \)

Lemma 7.3 (ADCB Lemma). Let \( \varepsilon, \Delta > 0 \). There exists \( L = L(S, \varepsilon, \Delta) > 0 \) and \( R_0 = R_0(S, \varepsilon, \Delta) > 0 \) with the following properties. Let \( A, B, C, D, T \in \pi_1(S, *) \) such that \( I(\cdot B \cdot) , I(\cdot D \cdot) , I(\cdot T \cdot) > L \). If \( R > R_0 \) and

\[
I(\cdot T \cdot A \cdot T \cdot) , I(\cdot T \cdot C \cdot T \cdot) , I(\cdot T \cdot B \cdot T \cdot) , I(\cdot T \cdot D \cdot T \cdot) \leq \Delta,
\]

then \( [ATBTCTDT] = [ATD\overline{T}CTB\overline{T}] \) in \( \Pi_{200\varepsilon,R} \) homology provided that the curves in question are in \( \Gamma_{\varepsilon,R} \).

Proof. Let \( \langle X, Y \rangle = [ATX^\ast CTY^\ast] \) for \( X, Y \in \pi_1(S, *) \), and let \( \{X, Y\} = \langle X, Y \rangle - \langle Y, X \rangle \) when both are in \( \Gamma_{\varepsilon,R} \). We claim that

\[
\{X, Y_0\} = \{X, Y_1\}
\]

in \( \Pi_{100\varepsilon,R} \) whenever \( I(TX\overline{T}) , I(TY\overline{T}) \leq \Delta \) and the curves in question are in \( \Gamma_{\varepsilon,R} \). To verify (11) we let \( A_i = Y_i \), \( B_0 = ATX\overline{T}C \), \( B_1 = CTX\overline{T}A \), and \( U = \overline{T} \) and \( V = T \), where \( A_i, B_i, U, V \) are from the statement of the Algebraic Square Lemma. Since by rotation

\[
\langle X, Y_0 \rangle = [Y_0 \overline{T}ATX\overline{T}CT]
\]

\[
\langle Y_0, X \rangle = [Y_0 \overline{T}CTX\overline{T}AT]
\]

\[
\langle X, Y_1 \rangle = [Y_1 \overline{T}ATX\overline{T}CT]
\]

\[
\langle Y_1, X \rangle = [Y_1 \overline{T}CTX\overline{T}AT],
\]
equation (11) follows from the Algebraic Square Lemma. (The hypotheses in the Algebraic Square Lemma follow from the hypothesis of this lemma and the Sum of Inefficiencies Lemma.) Likewise,

\[(12) \quad \{X_0, Y\} = \{X_1, Y\}\]

under the appropriate hypotheses.

In order to prove the lemma we first suppose that \(|l(TB\overline{T}) - l(TD\overline{T})| < \frac{\varepsilon}{4}\). If \(L\) is large enough (and hence \(l(B), l(D)\) and \(l(T)\) are large enough), it follows from the Connection Lemma that we can find a random geodesic arc

\[\cdot E \in \text{Conn}_{\varepsilon, l(T)}(\cdot T - 2l(T))(-i(T), i(T))\].

For any such \(E\), we have \(|l(TB\overline{T}) - l(TD\overline{T})| < \varepsilon + O(\varepsilon^2)\). Therefore, by the previous claim we have that the curves \(\langle B, E \rangle, \langle E, B \rangle, \langle D, E \rangle, \) and \(\langle E, D \rangle\) are in \(\Gamma_{2\varepsilon, R}\). Then from (11) and (12) it follows that

\[\{B, D\} = \{B, E\} = \{D, E\} = \{D, B\},\]

and therefore \(\{B, D\} = 0\), in \(\Pi_{200\varepsilon, R}\) homology.

More generally, if \(l(B), l(D) > L\), let \(k\) be the smallest integer such that

\[k > 4 \frac{|l(TD\overline{T}) - l(TB\overline{T})|}{\varepsilon}\]

Set

\[r_i = \frac{i}{2k} l(TD\overline{T}) + \frac{2k - i}{2k} l(TB\overline{T}) - 2l(T)\].

For \(0 < i < 2k\), we take random \(\cdot E_i \in \text{Conn}_{\varepsilon, r_i}(t(T)i(T))\) (observe that \(r_i > L - \Delta\)), and we let \(E_0 = D\) and \(E_{2k} = B\). Then

\[
\begin{align*}
\{E_0, E_{2k}\} &= \{E_1, E_{2k}\} \\
&= \{E_1, E_{2k-1}\} = \{E_2, E_{2k-1}\} \\
&= \{E_2, E_{2k-2}\} = \{E_3, E_{2k-2}\} \\
&\quad \vdots \\
&= \{E_{k-1}, E_{k+1}\} = \{E_k, E_{k+1}\}
\end{align*}
\]

and \(\{E_k, E_{k+1}\} = 0\) in \(\Pi_{200\varepsilon, R}\) homology by the first case, so we are finished.

Randomization (Randomization remarks for the ADCB Lemma). All constants \(K\) may depend only on \(\varepsilon, S\) and \(\Delta\). We have defined a map \(g : G^4 \to \mathbb{R}\Pi_{1, R}\) such that

\[\partial g(A, B, C, D) = [ATB\overline{T}CTD\overline{T}] - [ATD\overline{T}CTB\overline{T}]\]
In particular, we defined \( h : G^5 \to \mathbb{R} \Pi_{1,R} \) so that
\[
\partial h(A, C, X, Y_0, Y_1) = \{X, Y_0\} - \{X, Y_1\}.
\]
This map \( h \) is \( \varepsilon^{4l(T \cdot)} \) \( K \)-semirandom with respect to the measure classes \( \Sigma_G \times \Sigma_G \times \Sigma_G \times \Sigma_G \times \Pi \).

Then \( g(A, B, C, D) \) is a sum of \( 2k \) terms of the form \( \partial h(A, C, X, Y_0, Y_1) \), where each of \( X, Y_0, Y_1 \) is either \( B \) or \( D \), or \( E_i \), which is a \( K \)-semirandom element of \( G \) with respect to \( \Sigma_G \). Moreover, the \( Y_i \) are always independent from \( X \). Therefore, for each choice we make of \( X, Y_0, Y_1 \) (such as \( X = E_i \), \( Y_0 = E_{2k-i} \), \( Y_1 = E_{2k-i+1} \) or \( X = B \), and \( Y_0 = D \), \( Y_1 = E \)), the map from \( (A, B, C, D) \) to \( (A, C, X, Y_0, Y_1) \) is \( K \)-semirandom with respect to \( \Sigma_G^{\times 4} \) and \( \Sigma_G^{\times 2} \times \Sigma_G \times \Sigma_G^{\times 3} \). Therefore, noting that \( k < \frac{[8R]}{\varepsilon} \), we find that \( g \) is \( KRe^{4l(T \cdot \cdot \cdot)} \)-semirandom, with respect to \( \Sigma_G^{\times 4} \).

**Remark.** This remark is about the previous randomization remark. Where \( B \) and \( D \) are close in length, we can write \( \{B, D\} = \{B, B\} \) by (11), and hence \( \{B, D\} = 0 \). But we are letting \( (X, Y_0, Y_1) \) be \( (B, D, B) \), and the map \( (B, D) \to (B, D, B) \) is not 1-semirandom for \( \Sigma_G^{\times 2} \) and \( \Sigma_G \times \Sigma_G \times \Sigma_G^{\times 3} \) (because \( X \) and \( Y_1 \) are not independent). This map is only \( \varepsilon^{1(B \cdot \cdot \cdot \cdot \cdot)} \)-semirandom, which is no good. It is for this reason that we introduce \( E \).

The following lemma is a corollary of the *ADCB* Lemma. We call it the Four-part Itemization Lemma.

**Lemma 7.4** (Four-part Itemization Lemma). Let \( \varepsilon, \Delta > 0 \). There exists \( L = L(\varepsilon, \Delta) > 0 \) such that for any \( A, B, C, D, T \in \pi_1(S, \ast) \), we have
\[
[ATBTC\bar{T}\bar{C}TDT] - [\bar{T}\bar{D}T\bar{C}\bar{T}\bar{B}\bar{T}\bar{A}] = 2(A_T + B_T + C_T + D_T)
\]
in \( \Pi_{200\varepsilon, R} \) homology provided that \( I(\cdot \cdot \cdot T) > L \) and \( I(\cdot \cdot \cdot A \cdot \bar{T} \cdot B \cdot \bar{T} \cdot C \cdot \bar{T} \cdot D \cdot T) < \Delta \), and the curve \( [ATBTC\bar{T}\bar{C}TDT] \) is in \( \Gamma_{\varepsilon, R} \).

**Proof.** Recall the remark after the statement of Lemma 7.1. We have
\[
[ATBTC\bar{T}\bar{C}TDT] - [\bar{A}\bar{T}B\bar{T}C\bar{T}\bar{C}TDT] = 2A_T,
\]
\[
[\bar{A}\bar{T}B\bar{T}C\bar{T}\bar{C}TDT] - [\bar{A}\bar{T}\bar{B}T\bar{C}T\bar{D}TDT] = 2B_T,
\]
\[
[\bar{A}\bar{T}\bar{B}T\bar{C}\bar{T}\bar{D}TDT] - [\bar{A}\bar{T}\bar{B}T\bar{C}\bar{T}\bar{D}TDT] = 2C_T,
\]
\[
[ATBTC\bar{T}\bar{C}TDT] - [ATBTC\bar{T}\bar{C}TDT] = 2D_T
\]
in \( \Pi_{100\varepsilon, R} \) homology. (All the curves in question lie in \( \Gamma_{2\varepsilon, R} \) by the Flipping Lemma.) So
\[
[ATBTC\bar{T}\bar{C}TDT] - [ATBTC\bar{T}\bar{C}TDT] = 2(A_T + B_T + C_T + D_T)
\]
in \( \Pi_{100\varepsilon, R} \) homology. But
\[
[ATBTC\bar{T}\bar{C}TDT] - [ATBTC\bar{T}\bar{C}TDT] = 0
\]
in \( \Pi_{200\varepsilon, R} \) homology by the *ADCB* Lemma, so we are finished. \( \square \)
Randomization (Randomization remark for the Four-part Itemization Lemma). We have defined \( g : G^4 \to \mathbb{R}\mathfrak{P}_{1,R} \) such that
\[
\partial g(A, B, C, D) = [ATBT CD] - (A_T + B_T + C_T + D_T).
\]
This map is \( KRe_{\mathbb{A}}^{4\mathfrak{P}}\)-semirandom with respect to \( \Sigma \times 4 \mathfrak{P} \) and \( \sigma_{\mathfrak{P}} \) for some \( K = K(\varepsilon, S) \).

8. The \( XY \) Theorem

In this section we prove the \( XY \) Theorem, which states that
\[
(XY)_T = X_T + Y_T
\]
for suitable \( X, Y \) and \( T \).

The \( XY \) Theorem will be the central identity in the last section of the paper; it will allow us to reduce the encoding of long elements of \( \pi_1(S, \ast) \) to encoding of the generators. To prove the \( XY \) Theorem we will first prove two related statements called the First and the Second Rotation Lemmas. These are in turn proven with the Four-Part Itemization Lemma and the estimates from the Theory of Inefficiency.

8.1. The Rotation Lemmas. Let \( X, Y, Z \in \pi_1(S, \ast) \). Then we have the three geodesic arcs \( \cdot X \cdot, \cdot Y \cdot, \) and \( \cdot Z \cdot \). Consider the union of these three geodesic arcs as a \( \theta \)-graph on the surface \( S \). This \( \theta \)-graph generates an immersed pair of pants in \( S \) if and only if the triples of unit vectors \( i(\cdot X \cdot), i(\cdot Y \cdot), i(\cdot Z \cdot) \) and \( t(\cdot X \cdot), t(\cdot Y \cdot), t(\cdot Z \cdot) \) have the opposite cyclic orderings.

The following is the First Rotation Lemma.

**Lemma 8.1** (First Rotation Lemma). Let \( \varepsilon, \Delta > 0 \). There exists \( K = K(\varepsilon, \Delta) > 0 \) with the following properties. Let \( R_i, S_i, T \in \pi_1(S, \ast) \), \( i = 0, 1, 2 \), such that

(i) \( I(\cdot T \cdot R_i \cdot \bar{R}_{i+1} \cdot \bar{T} \cdot), I(\cdot T \cdot S_i \cdot \bar{S}_{i+1} \cdot \bar{T} \cdot) < \Delta \).

(ii) \( I(\cdot T \cdot \cdot) \geq K \).

(iii) \( I(\cdot R_i \cdot) + I(\cdot S_i \cdot) + 2I(\cdot T \cdot) < R - K \).

(iv) The triples of vectors \( (t(\cdot TR_i \cdot)) \) and \( (t(\cdot TS_i \cdot)) \), \( i = 0, 1, 2 \), have opposite cyclic ordering in \( T^*_i S \). (One of them is clockwise and the other one anti-clockwise.)

Then
\[
\sum_{i=0}^{2}(R_i \bar{R}_{i+1})_T + \sum_{i=0}^{2}(S_i \bar{S}_{i+1})_T = 0
\]

in \( \mathfrak{P}_{300\varepsilon,R} \) homology.
Remark. It follows by relabeling that if \( \max(I(\cdot R_i \cdot)) + \max(I(\cdot S_i \cdot)) + 2I(\cdot T \cdot) < R - K \) and the triples of vectors \((t(\cdot TR_i \cdot))\) and \((t(\cdot TS_i \cdot))\), \(i = 0, 1, 2\), have the same cyclic ordering in \( T_i S \), then

\[
\sum_{i=0}^{2} (R_i R_{i+1})_T + \sum_{i=0}^{2} (S_i S_{i+1})_T = 0
\]

in \( \Pi_{3\varepsilon,R} \) homology.

Proof. Let \( r_i \geq 0, i = 0, 1, 2 \), be the solutions of the equations

\[
r_i + r_{i+1} = 2R - I(\cdot TR_i \cdot) - I(\cdot TS_i \cdot) - I(\cdot T \cdot).
\]

Then we let \( A_i \) be a random element of \( \text{Conn}_{\varepsilon,r_i}(-i(\cdot T \cdot), i(\cdot T \cdot)) \).

Consider the three elements \( R_i \bar{T} A_i T S_i \) of \( \pi_1(S,*) \) and the corresponding geodesic arcs \( \bar{R}_i \bar{T} A_i T S_i \). We will show that the corresponding \( \theta \)-graph generates an immersed pair of pants \( \Pi_A \) in \( S \). The three cuffs of \( \Pi_A \) are the closed curves \([R_i+1 \bar{T} A_{i+1} T S_{i+1} S_i T A_i T R_i]\). We will also show that these closed geodesics have length \( 3\varepsilon \) close to \( 2R \), which implies that \( \Pi_A \in \Pi_{3\varepsilon,R} \).

We finish the argument as follows. Taking the boundary of \( \Pi_A \), we obtain

\[
\sum_{i=0}^{2} [R_i \bar{T} A_i T S_i \bar{S}_{i+1} \bar{T} A_{i+1} T R_{i+1}] = 0
\]
in $\Pi_{3\varepsilon,R}$ homology. Applying the Four-part Itemization Lemma we find

$$0 = \sum_{i=0}^{2} [R_{i+1}\bar{R}_i \bar{T}A_{i}TS_{i+1} \bar{T}\bar{A}_{i+1}T]$$

$$= \sum_{i=0}^{2}((R_{i+1}\bar{R}_i)T + (A_i)_T + (S_{i+1})T + (\bar{A}_{i+1})_T)$$

$$= \sum_{i=0}^{2}(R_{i}\bar{R}_{i+1})T + \sum_{i=0}^{2}(S_{i}\bar{S}_{i+1})T$$

in $\Pi_{300\varepsilon,R}$, because $(A_i)_T = -(\bar{A}_i)_T$.

We now verify that $[\bar{R}_i \bar{T}A_{i}TS_{i+1} \bar{T}\bar{A}_{i+1}T R_{i+1}] \in \Gamma_{2\varepsilon,R}$. By the New Angle Lemma (Lemma 4.2) applied to $\beta = \bar{T}$, and $\alpha = \bar{T} \cdot R_i\bar{\cdot}$, for $K$ large enough (and therefore $I(\cdot \cdot \cdot)$ large), the angle $\Theta(i(\cdot \cdot \cdot), i(\cdot \cdot \cdot)) \leq \frac{\varepsilon}{10}$, and likewise $\Theta(t(\cdot \cdot \cdot), t(\cdot \cdot \cdot)) \leq \frac{\varepsilon}{10}$, and for the same with $R_i$ replaced with $S_i$. It follows that $\Theta(i(\cdot \cdot \cdot), i(\cdot \cdot \cdot)) < 2\varepsilon$, and so on. So by the Sum of Inefficiencies Lemma (using equation (5)),

$$|I[\bar{R}_i \bar{T}A_{i}TS_{i+1} \bar{T}\bar{A}_{i+1}T R_{i+1}]) - I(\cdot A_i \cdot) - I(\cdot TR_{i+1}\bar{R}_i \bar{T} \cdot) - I(\cdot A_{i+1} \cdot)$$

$$- I(\cdot TS_{i} \bar{S}_{i+1} \bar{T} \cdot) | < O(\varepsilon^2),$$

and moreover by (15), we have

$$|I(\cdot A_i \cdot) + I(\cdot TR_{i+1}\bar{R}_i \bar{T} \cdot) + I(\cdot A_{i+1} \cdot) + I(\cdot TS_{i} \bar{S}_{i+1} \bar{T} \cdot) - 2R| < 2\varepsilon,$$

which proves the claim.

We now verify that the $\theta$-graph associated to the geodesic arcs $\bar{R}_i \bar{T}A_{i}TS_{i} \cdot$ generates an immersed pair of pants $\Pi_A$ in $S$. We find the unique $\theta_0 \in [0, \pi]$ such that $I(\pi - \theta_0) = \Delta + 1$ (so $\pi - \theta_0 = 2 \sec^{-1}(e^{-\frac{\Delta+1}{2}})$). Observe that $I(\cdot \bar{R}_i \bar{T} \cdot) < I(\cdot T \cdot R_{i+1} \cdot \bar{R}_i \bar{T} \cdot) \leq \Delta$. Then $I(\cdot R_{i} \bar{T} \cdot) > 1(\cdot T \cdot) - \Delta > K - \Delta$. By the Sum of Inefficiencies Lemma for Angles,

$$I(\pi - \Theta(i(\cdot \bar{R}_i \bar{T} \cdot), i(\cdot \bar{R}_{i+1} \bar{T} \cdot))) - 1 < I(\cdot TR_{i+1} \cdot \bar{R}_i \bar{T} \cdot)$$

$$\leq I(\cdot T \cdot R_{i+1} \cdot \bar{R}_i \bar{T} \cdot) \leq \Delta.$$

Therefore, $\Theta(i(\cdot \bar{R}_i \bar{T} \cdot), i(\cdot \bar{R}_{i+1} \bar{T} \cdot)) > \theta_0$.

On the other hand, by the New Angle Lemma, because the geodesic arc $\cdot \bar{R}_i \bar{T} \cdot$ is long for large enough $K$ (we showed above that $I(\cdot \bar{R}_i \bar{T} \cdot) > K - \Delta$), we have $\Theta(i(\cdot \bar{R}_i \bar{T} \cdot), i(\cdot \bar{R}_i \bar{T}A_{i}TS_{i} \cdot)) < \frac{\theta_0}{2}$, so the cyclic order of the triple of vectors $i(\cdot \bar{R}_i \bar{T}A_{i}TS_{i} \cdot)$, $i = 0, 1, 2$, is the same as of the triple of vectors $i(\cdot \bar{R}_i \bar{T} \cdot)$, and likewise the cyclic order of the triple of vectors $t(\cdot \bar{R}_i \bar{T}A_{i}TS_{i} \cdot)$, $i = 0, 1, 2$, is the same as of the triple of vectors $t(\cdot TS_{i} \cdot)$. So the corresponding cyclic orderings are opposed and we are finished. $\square$
**Randomization** (Randomization remark for the First Rotation Lemma).
We let $K = K(\varepsilon, S)$. We have defined $g : G^6 \to \mathbb{R} \Pi_1, R$ such that

$$
\partial g(R_0, R_1, R_2, S_0, S_1, S_2) = 2 \sum_{i=0}^2 (R_{i+1} \bar{R}_i)_T + (S_i \bar{S}_{i+1})_T.
$$

Let $\Pi$ denote the pants whose $\theta$-graph is made out of the three connections $\cdot \bar{R}_i T A_i T S_i \bar{S}_{i+1} T A_{i+1} T R_{i+1} \cdot$, $i = 0, 1, 2$. We can write

$$
g((R_i), (S_i)) = \Pi + \sum_{i=0}^2 g_1(R_{i+1} \bar{R}_i, A_i, S_i \bar{S}_{i+1}, \bar{A}_{i+1}),
$$

where $g_1$ is the map from the Four-part-Itemization Lemma (see the randomization remark). So $g$ is $K(e^{12\varepsilon \cdot T \cdot} + Re^{4\varepsilon \cdot T \cdot})$-semirandom with respect to $\Sigma_G^6$ and $\Sigma_\Pi$.

The Second Rotation Lemma is

**Lemma 8.2** (Second Rotation Lemma). Let $\varepsilon, \Delta > 0$. There exists $K = K(\varepsilon, \Delta) > 0$ with the following properties. Let $R_i, T \in \pi_1(S, \cdot)$, $i = 0, 1, 2$, such that

(i) $I(\cdot \cdot T \cdot R_i \cdot R_{i+1} \cdot T \cdot) < \Delta$,
(ii) $I(\cdot \cdot T \cdot) \geq K$.

Then

$$
\sum_{i=0}^2 (R_i \bar{R}_{i+1})_T = 0
$$

in $\Pi_{300\varepsilon, R}$ homology.

**Proof.** Given $T$, we choose $v \in T_1^1 S$ and let $\rho = e^{2\pi i}$. We take $L$ sufficiently large so that $Conn_{\varepsilon, L}(t(\cdot \cdot T \cdot), \rho^i v)$ is nonempty for $i = 0, 1, 2$. Then we choose $\cdot S_i \cdot \in Conn_{\varepsilon, L}(t(\cdot \cdot T \cdot), \rho^i v)$. Then $I(\cdot \cdot T \cdot S_i \cdot \bar{S}_{i+1} T \cdot) \leq \log \frac{4}{3} + O(\varepsilon) \leq 1$, by the Sum of Inefficiencies for Angles Lemma, so when $I(\cdot \cdot T \cdot)$ is large we can apply the previous Lemma (see the remark after Lemma 8.1) with $R_i := S_i$ to obtain

$$
2 \sum_{i=0}^2 (S_i \bar{S}_{i+1})_T = 0
$$

in $\Pi_{300\varepsilon, R}$ homology.

Then given $R_i$ as in the hypothesis to this lemma, we obtain

$$
\sum_{i=0}^2 (S_i \bar{S}_{i+1})_T + (R_i \bar{R}_{i+1})_T = 0,
$$

and so

$$
\sum_{i=0}^2 (R_i \bar{R}_{i+1})_T = 0.
$$

\[\square\]
**Randomization** (Randomization remark for the Second Rotation Lemma). All constants $K$ may only depend on $\varepsilon$ and $S$. We have defined $g : G^3 \to \mathbb{R}\Pi_1, R$ such that

$$\partial g(R_0, R_1, R_2) = \sum_{i=0}^{2} (R_i R_{i+1})_T.$$

We are fixing $S_0, S_1, S_2$ of length $L$, so the triple $(S_0, S_1, S_2)$ is $e^{3L}$-semi-random, and the maps

$$(R_0, R_1, R_2) \to (R_0, R_1, R_2, S_0, S_1, S_2)$$

and

$$(R_0, R_1, R_2) \to (S_0, S_1, S_2, S_0, S_1, S_2)$$

are $e^{3L}$ and $e^{6L}$ semi-random respectively.

Then letting $g_{1R}$ be the $g$ for the First Rotation Lemma, we can letter

$$g(R_0, R_1, R_2) = g_{1R}((R_i), (S_i)) - \frac{1}{2} g_{1R}((S_i), (S_i))$$

and $g_{1R}$ is $K Re^{12t(-T)}$ semirandom, so $g$ is $K Re^{6L+12t(-T)}$ semirandom.

**8.2. The XY Theorem.** The following theorem follows from the Second Rotation Lemma. We call it the XY Theorem.

**Theorem 8.1 (XY Theorem).** Let $\varepsilon, \Delta > 0$. There exists $K = K(\varepsilon, \Delta) > 0$ with the following properties. Let $X, Y, T \in \pi_1(S, \ast)$, $i = 0, 1, 2$, such that

(i) $I(T \cdot X \cdot Y \cdot T \cdot)$, $I(T \cdot X \cdot \bar{T} \cdot)$, $I(T \cdot Y \cdot \bar{T} \cdot) < \Delta$;

(ii) $I(T \cdot \cdot) \geq K$.

Then $(XY)_T = X_T + Y_T$ in $\mathbb{H}_{300\varepsilon, R}$ homology.

**Proof.** Set $R_0 = \text{id}$, $R_1 = X$, and $R_2 = Y$, and apply the previous lemma. \hfill \Box

**Randomization.** We have defined the map $g_{XY} : (X, Y) \to \mathbb{R}\Pi_1, R$ such that $\partial g_{XY}(X, Y) = (XY)_T - X_T - Y_T$. (The map $g_{XY}$ is defined on the appropriate subset of $\Sigma^2_G$ described in the statement of Theorem 8.1.) This map is $K Re^{12t(-T)}$-semi-random with respect to $\Sigma^2_G$ and $\sigma_{\Pi}$, where $K = K(\varepsilon, \Delta)$.

**9. The endgame**

In this last section of the main text of the paper, we prove that every good curve is good pants homologous to a sum of encodings of a given set of standard generators for $\pi_1(S, \ast)$.

We prove this in three steps:

1. We prove in Lemma 9.2 that every good curve is good pants homologous to a sum of two encodings of two elements of $\pi_1(S, \ast)$; these elements are represented by geodesic segments of length about $R$. 
2. We use the XY Theorem and Lemma 9.3 to repeatedly write \( X_T = (X_1)_T + (X_2)_T \) where \( X_1 \) and \( X_2 \) have length about half that of \( X \). This allows us to reduce an encoding of an arbitrary element of \( \pi_1(S,*) \) to a sum of encodings of elements of bounded length (bounded in terms of \( S \) and \( \varepsilon \)).

3. We use the XY Theorem to reduce the encoding of an element of \( \pi_1(S,*) \) of bounded length to a sum of encodings of generators. This requires the proper choice of \( T \), which is discussed in Lemma 9.1.

9.1. The good pants homology of short words. The following is the Good Direction Lemma.

**Lemma 9.1 (Good Direction Lemma).** For any finite set \( W \subset \pi_1(S,*) \), we can find \( \Delta = \Delta(S,W) \) such that for any \( L \), we can find \( T \in \pi_1(S,*) \) such that \( I(\cdot,T) \cdot L > 1 \) and \( I(\cdot,T \cdot X \cdot T^{-1}) < \Delta \), when \( X \in W \).

**Proof.** For any \( v \in T^1_1 S \) and \( t > 0 \), we let \( \alpha_t(v) \) be the geodesic segment of length \( t \) such that \( i(\alpha_t(v)) = v \), and we let \( \alpha_\infty(v) \) be the corresponding infinite geodesic ray. We claim that for any \( X \in \pi_1(S,*) \) and \( X \neq id \), there are at most two \( v \in T^1_1 S \) such that

\[
\lim_{t \to \infty} I(\alpha_t^{-1}(v) \cdot X \cdot \alpha_t(v)) = \infty.
\]

To prove the claim we lift \( X \cdot T \) to the universal cover \( \mathbb{H}^2 \), and thus get two lifts of \( T \), and hence two lifts of \( v \). We observe that (17) holds if and only if the two lifts of \( \alpha_\infty(v) \) end at the same point of \( \partial \mathbb{H}^2 \). The map that maps one lift of \( \alpha_\infty(v) \) to the other is the deck transformation that maps one lift of \( T \) to the other. The relation (17) holds if and only if \( \alpha_\infty(v) \) is a fixed point of the Möbius transformation \( M \), and since \( M \) is not the identity this can be true for at most two vectors \( v \).

In the rest of this section we fix a set of standard generators \( g_1, \ldots, g_{2n} \) of \( \pi_1(S,*) \). (Here \( n \) is the genus of \( S \).) Recall that \( H_1 \) denotes the standard homology on \( S \). Let \([g_i]\) denote the corresponding closed curves. For any closed curve \( \gamma \in S \), there are unique \( a_1, \ldots, a_{2n} \) such that \( \gamma = \sum a_i[g_i] \) in \( H_1 \). We define \( q : \Gamma \to \mathbb{R} \pi_1(S,*) \) by \( q(\gamma) = \sum a_i g_i \), where \( \Gamma \) is the set of all closed curves on \( S \). We extend the definition of \( q \) to a map \( q : \pi_1(S,*) \to \mathbb{R} \{g_1, \ldots, g_{2n}\} \) by \( q(X) = q([X]) \).

For \( l \in \mathbb{N} \), we define the set \( W_l \) as the set of elements \( X \in \pi_1(S,*) \) that can be written as a product of at most \( l \) generators (or their inverses).

**Theorem 9.1.** Let \( \varepsilon > 0 \). For all \( l \in \mathbb{N} \) and \( L > 0 \), we can find \( T \in \pi_1(S,*) \) and \( R_0 \) such that \( I(\cdot,T) > L \), and for \( R > R_0 \) and \( X \in W_l \), we have

\[
X_T = (q(X))_T
\]

in \( H_{300e,R} \) homology.
Remark. Here we extended the partial map \((\cdot)_T : \pi_1(S, *) \to \mathbb{R} \Gamma_{\varepsilon, R}\) (given by \(X \mapsto X_T\)) to a partial map \((\cdot)_T : \mathbb{R} \pi_1(S, *) \to \mathbb{R} \Gamma_{\varepsilon, R}\). We remind the reader that \(X_T\) depends implicitly on \(R\) and \(\varepsilon\).

Proof. We take \(\Delta = \Delta(W_i)\) and \(T = T(W_i, L)\) from the previous lemma, so \(1(T \cdot X) > L\) and \(I(T \cdot X \cdot T_1) < \Delta\) for all \(X \in W_i\). If \(X \in W_i\), then \(q(X) = X\) or \(q(X) = -X\), so \(X_T = (q(X))_T\).

Take \(1 \leq k < l\), and assume \(X_T = ((q(X))_T)\) in the \(\Pi_{300\varepsilon, R}\) homology for all \(X \in W_k\). Then for any \(X \in W_{k+1}\), we can write \(X = g_{\varepsilon}Y\), for some \(i \in \{1, \ldots, 2n\}\), and \(\sigma = \pm 1\), and \(Y \in W_k\). Then \(X_T = (g_{\varepsilon}^\gamma)_{T_T} + Y_T\) by the \(XY\) Theorem (see Theorem 8.1), which requires

\[I(T \cdot X \cdot T_1), I(T \cdot g_{\varepsilon}^\gamma \cdot T_1), I(T \cdot Y \cdot T_1) < \Delta,\]

and \(Y_T = (q(Y))_T\) by assumption, so \(X_T = ((q(X))_T\). We conclude the theorem by induction. \(\square\)

Randomization (Randomization remarks for Theorem 9.1). Given \(l, L, T\) and \(R\) (and \(\varepsilon\)), we have implicitly defined the map \(g_W : W_l \to \Pi_{300\varepsilon, R}\) such that \(\partial g_W(X) = X_T - (q(X))_T\). The map \(g_W\) arises from a sum of at most \(l\) applications of the \(XY\) Theorem, so \(g_W\) is \(K(S)RK\varepsilon^{12l(T \cdot \cdot)}\)-semirandom, because every measure in \(\Sigma_{\varepsilon}\) has total mass at most \(K(S)\).

9.2. Preliminary lemmas. We now observe that every good curve is good pants homologous to \((X_0)_T + (X_1)_T\) for suitable \(X_0\) and \(X_1\) from \(\pi_1(S, *)\).

Lemma 9.2. There exists a universal constant \(\bar{\varepsilon} > 0\) such that for every \(0 < \varepsilon < \bar{\varepsilon}\), there exist constants \(L = L(\varepsilon, S) > 0\) and \(R_0 = R_0(\varepsilon, S) > 0\) with the following properties. For any \(i \in \Gamma_{\varepsilon, R}\) and \(T \in \pi_1(S, *)\), \(1(T \cdot X) > L\), we can find \(X_0, X_1 \in \pi_1(S, *)\) such that

(i) \(|\partial \cdot X_i \cdot ) - (R + 2L - \log 4)| < \frac{1}{2};
(ii) \(\Theta (t \cdot X_i \cdot ) = \Theta (t \cdot X_i \cdot ) \leq -\frac{\pi}{6};
(iii) \(\gamma = (X_0)_T + (X_1)_T\) in \(\Pi_{300\varepsilon, R}\) homology for \(R > R_0\).

Proof. We take at random two points \(x_0\) and \(x_1\) on the parametrizing torus \(T_\gamma\) that are \(\text{hl}(\gamma)\) apart, and we let \(w_i \in T_{x_i}^1\) be \(-\sqrt{-1}T_\gamma'(x_i)\). We let \(\gamma_i\) be the subsegment of \(\gamma\) from \(x_i\) to \(x_{i+1}\) (where \(x_1 = x_0\)).

For \(i = 0, 1\), we take \(\alpha_i \in \text{Conn}_{\bar{\varepsilon}, L}(t \cdot T_1, w_i)\), where \(L = L(\varepsilon, S)\) is the constant from the Connection Lemma. (That is, we choose \(L\) so that the set \(\text{Conn}_{\varepsilon, L}(t \cdot T_1, w_i)\) is nonempty.) Observe that the piecewise geodesic arc \(\alpha_0 \alpha_1 \alpha_1^{-1}\) begins and ends at the point *, so we let \(X_0 \in \pi_1(S, *)\) denote the corresponding element of \(\pi_1(S, *)\). Similarly, we let \(X_1 \in \pi_1(S, *)\) be the element that corresponds to the curve \(\alpha_1 \gamma_1 \alpha_0^{-1}\).
It follows from the remark after Lemma 4.8 that inequality (i) of the statement of the lemma holds. On the other hand, by the New Angle Lemma the angle \( \Theta(i(\cdot X_0 \cdot), i(\alpha_0)) \) is as small as we want providing that \( I(\alpha_0) > L \) is large enough. (Here we use that the inefficiency \( I(\alpha_0 \gamma_0 \gamma_1^{-1}) \) is bounded above.) Since by construction the angle \( \Theta(i(\alpha_0), i(\cdot T \cdot)) \) is less than \( \frac{\pi}{10} \), we conclude that for \( L \) large enough, we have \( \Theta(t(\cdot T \cdot), i(\cdot X_0 \cdot)) < \frac{\pi}{6} \). Other cases are treated similarly.

Let \( \cdot A \cdot \) be a random element of \( \text{Conn}_{\text{m}} R((\cdot T \cdot), \cdot (\cdot T \cdot)) \), where \( R' = R + 2L - 2(\cdot T \cdot) \). Then

\[
|I([X_0 T A T]) - 2R| < \varepsilon, \quad |I([X_1 T A T]) - 2R| < \varepsilon,
\]

so \( \gamma = [X_0 T A T] + [X_1 T A T] \) in \( \Pi_{\varepsilon, R} \) homology, because the three curves bound a good pair of pants.

Moreover, \( [X_0 T A T] = (X_0)_T + (\tilde{A})_T \), and \( [X_1 T A T] = (X_1)_T + A_T \) in \( \Pi_{100\varepsilon, R} \) homology by the Two-part Itemization Lemma. Since \( (\tilde{A})_T = -A_T \) we conclude \( \gamma = (X_0)_T + (X_1)_T \) in \( \Pi_{300\varepsilon, R} \) homology.

**Randomization** (Randomization remarks for Lemma 9.2). We have defined the maps \( q_C : \Gamma_{1, R} \to \mathbb{R} G \) (by \( q_C(\gamma) = X_0 + X_1 \)) and \( g_C : \Gamma_{1, R} \to \mathbb{R} \Pi_{1, R} \), such that \( \partial g_C(\gamma) = \gamma - (g_C(\gamma))_T \) (where \( A \mapsto A_T \) maps \( \mathbb{R} G \to \mathbb{R} \Gamma_{1, R} \)).

The map \( g_C \) is \( 3^L K(-) \) K-semirandom with respect to \( \sigma_T \) and \( \Gamma_T \). The map \( g_C \) is \( 3^{2L} K(S, \varepsilon) \) semirandom with respect to \( \sigma_{\Pi_{1}} \) and \( \sigma_{\Pi_{1}} \), where \( K = K(S, \varepsilon) \).

We have the following definition. For any \( X, T \in \pi_1(S, \ast) \), \( X \not\equiv \ast \), we let

\[
\theta^T_X = \max\{\Theta(\cdot T \cdot, i(\cdot T \cdot)), \Theta(\cdot T \cdot, i(\cdot T \cdot))\}.
\]

**Lemma 9.3.** For \( L > L_0(S) \) and \( X, T \in \pi_1(S, \ast) \), \( X \not\equiv \ast \), then we can write \( X = X_0 X_1 \), for some \( X_0, X_1 \in \pi_1(S, \ast) \), such that

(i) \( |I(1; \cdot X_1 \cdot) - \left(\frac{L}{2} - L - \log 2\right)| < \frac{1}{2} \);

(ii) \( I(\cdot X_0, \cdot X_1 \cdot) \leq 2L + 3 \);

(iii) \( \theta^T_{X_1} \leq \max\{\theta^T_X + e^{L+4} e^{-I(\cdot X_1 \cdot)}, \frac{\pi}{6}\} \).

**Proof.** We let \( \alpha = \cdot X \cdot \). Then \( \alpha : [0, I(\cdot X \cdot)] \to S \) is the unit speed parametrization with \( \alpha(0) = \alpha(I(\cdot X \cdot)) = \ast \). We let \( y = \frac{I(\cdot X \cdot)}{2} \). Then for \( L \) large enough, we can find \( \beta \in \text{Conn}_{\text{m}} L(t(\cdot T \cdot), \sqrt{-1} \alpha' (y)) \). (As always, \( L \) is determined by the Connection Lemma.)

Then \( \alpha[0, y] \beta^{-1} \) begins and ends at \( \ast \), so it represents some \( X_0 \in \pi_1(S, \ast) \). Likewise \( \beta \alpha[y, I(\cdot X \cdot)] \) represents some \( X_1 \in \pi_1(S, \ast) \), and \( X = X_0 X_1 \). Moreover, it follows from the remark after Lemma 4.8 that

\[
I(\cdot X_1 \cdot) - \left(\frac{L}{2} - L - \log 2\right) < \frac{1}{2}.
\]

Condition (ii) follows immediately from (i).
Let \( \theta = \Theta(i(\cdot X\cdot), i(\cdot X_0\cdot)) \). Then by the hyperbolic law of sines, assuming that \( I(\cdot X\cdot) \geq 1 \) (which follows if we assume that \( I(\alpha) \geq L-1 \) is at least 1), we obtain
\[
\sin(\theta) \leq \frac{\sinh(L+1)}{\sinh(I(\cdot X_0\cdot))} \leq e^{L+2-I(\cdot X_0\cdot)}.
\]
Therefore,
\[
\Theta(t(\cdot T\cdot), i(\cdot X_0\cdot)) \leq \Theta(t(\cdot T\cdot), i(\cdot X\cdot)) + e^{L+4-I(\cdot X_0\cdot)}.
\]

By similar reasoning we find that \( \Theta(t(\cdot X_0\cdot), -i(\beta)) \leq e^{2-L} < \frac{\pi}{12} \), assuming that \( L \) is large enough. Also by construction, \( \Theta(-i(\beta), t(\cdot T\cdot)) < \frac{1}{20} < \frac{\pi}{12} \), so \( \Theta(t(\cdot X_0\cdot), i(\cdot T\cdot)) \leq \frac{\pi}{7} \). We proceed similarly for \( X_1 \).

**Randomization** (Randomization remarks for Lemma 9.3). We have defined \( \tilde{q}_D : G \to G^2 \) such that \( \tilde{q}_D(X) = (X_0, X_1) \). If \( I(\cdot X\cdot) \in [a, a+1] \), then \( I(\cdot X_0\cdot), I(\cdot X_1\cdot) \in \left[ \frac{a}{2} + L', \frac{a}{2} + L' + 1 \right] \), where \( L' = L - \log 2 - \frac{1}{2} \).

Moreover, given \((X_0, X_1) \in G^2\), there is at most one \( X \) such that \( \tilde{q}_D(X) = (X_0, X_1) \) (because \( X = X_0X_1 \)). We conclude that
\[
(\tilde{q}_D)_*\sigma_0 \leq e^{2L'+2}\sigma_{\frac{a}{2}+L'} \times \sigma_{\frac{a}{2}+L'},
\]
and hence \( \tilde{q}_D \) is \( e^{2L'+2} \)-semirandom. It follows that the map \( q_D : G \to \mathbb{R}G \) defined by \( X \to X_0 + X_1 \) is \( 2e^{2L'+2} \)-semirandom.

**9.3. Proof of Theorem 3.3.** The following theorem implies Theorem 3.3. Recall that \( \{g_1, \ldots, g_{2n}\} \) denotes a standard basis for \( \pi_1(S, *) \), where \( n \) is the genus of \( S \).

**Theorem 9.2.** Let \( \varepsilon > 0 \). There exists \( R_0 = R_0(S, \varepsilon) > 0 \) with the following properties. There exists \( T \in \pi_1(S, *) \), where \( T \) depends only on \( \varepsilon \) and \( S \), such that for every \( R > R_0 \) and every \( \gamma \in \Gamma_{\varepsilon, R} \), we have
\[
\gamma = \sum_{i=1}^{2g} a_i(g_i)_T
\]
in \( \Pi_{300\varepsilon, R} \) homology for some \( a_i \in \mathbb{Q} \).

**Remark.** To prove Theorem 3.3 we take \( h_i = (g_i)_T \). Since \( (g_i)_T \) is equal to the closed curve on \( S \) that corresponds to \( g_i \) in the standard homology \( H_1 \), it follows that \( h_i \) is a basis for \( H_1 \) (with rational coefficients).

**Proof.** We take \( L \) that is sufficiently large for Lemmas 9.2 and 9.3. We let \( l \in \mathbb{N} \) be such that \( X \in W_l \) whenever \( I(\cdot X\cdot) < 2L + 5 \). Then by Theorem 9.1 we can find \( T \) such that \( I(\cdot T\cdot) > L \) and \( I(\cdot T\cdot) > K(\varepsilon, 2L + 3) \), where \( K(\varepsilon, \Delta) \) is the constant from Theorem 8.1, and such that \( X_T = (q(X))_T \) in \( \Pi_{300\varepsilon, R} \) homology for all \( X \in W_l \). We take \( R > R_0(S, \varepsilon, L) \) from Lemma 9.2, and \( R > R_0(L, T) \) from Theorem 9.1.
Fix any $\gamma \in \Gamma_{\epsilon,R}$. By Lemma 9.2 we can find $X_0, X_1 \in \pi_1(S, \ast)$ such that

$$|I(\cdot | X_i \cdot) - (R + 2L - \log 4)| < \frac{1}{2}$$

and

$$(18) \quad \gamma = (X_0)_T + (X_1)_T$$

in $\Pi_{300\epsilon,R}$ homology. Observe that $q(\gamma) = q(X_0) + q(X_1)$.

By Lemma 9.3 we can write

$$X_0 = X_{00} X_{01},$$

where

$$(19) \quad l(\cdot | X_{0i} \cdot) \in \mathbb{R}^2 + 2L, \mathbb{R}^2 + 2L + 1\,$$

and the conclusions of Lemma 9.3 hold. And likewise for $X_1$.

Let $N = \lfloor \log_2 R \rfloor - 1$. For every $0 \leq k \leq N$, we define sets $X_k$ by letting $X_0 = \{X_0, X_1\}$ and the set $X_{k+1}$ is the set of children of elements of $X_k$. Each set $X_k$ has $2^{k+1}$ elements and the elements of $X_k$ are not necessarily distinct. (For the pedantic reader, we proceed as follows: $X_k$ is a set of ordered pairs of the form $(a, X)$, when $0 \leq a < 2^k$ and $X \in \pi_1(S, \ast)$. If $X_0, X_1$ are constructed from $X$ according to Lemma 9.3, we let the children of $(a, X)$ be $(2a + i, X_i)$ for $i = 0, 1$. Then we let $X_0 = \{(i, X_i) : i = 0, 1\}$ and let $X_{k+1}$ be the set of children of $X_i$.)

Moreover, for any $X \in X_k$, we have

$$I(\cdot | X \cdot) \in [R 2^{-k} + 2L, R 2^{-k} + 2L + 1].$$

We claim that

$$\theta^T_X < \frac{\pi}{3}$$

for every $X$ in any $X_k$. For any such $X$, we can find a sequence $Y_0, Y_1, \ldots, Y_k$, so that $Y_0 = X_0$ or $Y_0 = X_1$ and $Y_k = X$, and where $Y_{i+1}$ is a child of $Y_i$. It follows from equation (19) that $I(\cdot | Y_{i+1} \cdot) \leq I(\cdot | Y_i \cdot) - 1$, and $I(\cdot | Y_k \cdot) \geq 2L - 2$, and

$$\theta^T_{Y_k} \leq \frac{\pi}{6} + \sum_{i=0}^{k} e^{L+4-I(\cdot | Y_i \cdot)}$$

$$\leq \frac{\pi}{6} + \frac{e^{L+4-(2L-3)}}{e-1} < \frac{\pi}{3},$$

assuming $L > 8$.

By Lemma 4.3 we have $I(\cdot | T \cdot X \cdot \vec{T} \cdot) \leq \log 4$ for every $X$ in every $X_k$. Therefore, we can apply The $XY$ Theorem (see Theorem 8.1) and conclude that

$$(20) \quad Y_T = (Y_0 Y_1)_T$$

whenever $Y$ is a nontrivial node of our tree and $Y_0$ and $Y_1$ are its two children.
It follows from (18) and (20) applied recursively that
\[ \gamma = \sum_{X \in X_N} X_T \]
in \( \Pi_{300e,R} \) homology. We know that if \( X \in X_N \), then \( X \notin W_1 \), so \( X_T = (q(X))_T \). Therefore
\[ \gamma = \sum_{X \in X_N} (q(X))_T = \sum ((q(\gamma))_T, \] and so we are finished. \( \square \)

Randomization (Randomization remarks for the proof of Theorem 3.4). We have determined \( T \equiv T(S, \varepsilon) \), so \( e^{\mu(T, \cdot)} = K(S, \varepsilon) \). We have implicitly defined the map \( g : \Gamma_{\varepsilon,R} \rightarrow \mathbb{R} \Pi_{300e,R} \) such that \( \partial g(\gamma) = \gamma - (q(\gamma))_T \). We note that \( q = q_N \circ q_C \), where \( q_C(\gamma) = X + X' \) from Lemma 9.2 and \( q_D(X) = X_0 + X_1 \) from Lemma 9.3.

Moreover,
\[ g(\gamma) = g_C(\gamma) + \sum_{i=0}^{N-1} g_{XY}(q_D(q_D(q_C(\gamma)))) + g_W(q_N(q_C(\gamma))), \]
where \( g_C \) is the map from Lemma 9.2, \( q_D \) and \( q_D \) are the maps from Lemma 9.3, \( g_{XY} \) is the map from Theorem 8.1, \( N \) is the number of times we iterate the division (the application of Lemma 9.3), and \( g_W \) is from Theorem 9.1.

By far the most important point is that \( q_D \) is \( K = K(S, \varepsilon) \)-semirandom, so \( q_D^i \) is \( K^i \)-semirandom, for any \( i \leq N \) (recall that \( N \leq \lceil \log_2 R \rceil \)), and therefore \( K^i \leq R^{\log_2 K^i} \), so the map \( q_D \) is \( P(R) \)-semirandom, where \( P(R) \) denotes a polynomial in \( R \).

9.4. The proof of Theorem 3.4. The map \( \phi \) from Theorem 3.4 is defined to be equal to the map \( g \) from the Randomization remarks for Theorem 9.2. We take \( h_i = (g_i)_T \). Then \( \partial \phi(\gamma) = \gamma - (q(\gamma))_T \) and \( (q(\gamma))_T \in \mathbb{R}\{h_1, \ldots, h_2n\} \), for any \( \gamma \in \mathbb{R}\Gamma_{\varepsilon,R} \). Moreover, the map \( \phi \) is \( P(R) \)-semirandom as shown in those Randomization remarks. This implies estimate (iii) of the statement of Theorem 3.4, and we are finished.

10. Appendix 1

Introduction to randomization. Let \((X, \mu)\) and \((Y, \nu)\) denote two measure spaces (where \( \mu \) and \( \nu \) are positive measures).

Definition 10.1. We say that a map \( g : (X, \mu) \rightarrow (Y, \nu) \) is \( K \)-semirandom with respect to \( \mu \) and \( \nu \) if \( g_* \mu \leq K \nu \).

By \( \mathbb{R}X \) we denote the vector space of finite formal sums (with real coefficients) of points in \( X \). There is a natural inclusion map \( \iota : X \rightarrow \mathbb{R}X \), where \( \iota(x) \in \mathbb{R}X \) represents the corresponding sum. Then every map \( f : \mathbb{R}X \rightarrow S \), where \( S \) is any set, induces the map \( f : X \rightarrow S \) by letting \( f = \tilde{f} \circ \iota \).
Let $f : X \to \mathbb{R}Y$ be a map. Then we can write $f(x) = \sum_y f_x(y)y$, where the function $f_x : Y \to \mathbb{R}$ is nonzero for at most finitely many points of $Y$. We define $|f| : X \to \mathbb{R}Y$ by

$$|f|(x) = \sum_y |f_x(y)|y.$$  

We define the measure $|f|_*\mu$ on $Y$ by

$$|f|_*\mu(V) = \int_X \left( \sum_y |f_x(y)|\chi_V(y) \right) d\mu(x),$$

for any measurable set $V \subset Y$, and $\chi_V(y) = 1$ if $y \in V$, and $\chi_V(y) = 0$ if $y \notin V$.

**Definition 10.2.** Let $(X,\mu)$ and $(Y,\nu)$ be two measure spaces (with positive measures $\mu$ and $\nu$). A map $f : X \to \mathbb{R}Y$ is $K$-semirandom if $|f|_*\mu \leq K\nu$. A linear map $\tilde{f} : \mathbb{R}X \to \mathbb{R}Y$ is $K$-semirandom with respect to measures $\mu$ and $\nu$ on $X$ and $Y$ respectively if the induced map $f : X \to \mathbb{R}Y$ is $K$-semirandom.

The following propositions are elementary.

**Proposition 10.1.** Let $X$, $Y$ and $Z$ denote three measure spaces. If $f : \mathbb{R}X \to \mathbb{R}Y$ is $K$-semirandom, and $f : \mathbb{R}Y \to \mathbb{R}Z$ is $L$-semirandom, then $g \circ f : \mathbb{R}X \to \mathbb{R}Z$ is $KL$-semirandom.

**Proposition 10.2.** If $f_i : \mathbb{R}X \to \mathbb{R}Y$ is $K_i$-semirandom, $i = 1, 2$, and $\lambda_i \in \mathbb{R}$, then the map $\left(\lambda_1 f_1 + \lambda_2 f_2\right) : \mathbb{R}X \to \mathbb{R}Y$ is $(|\lambda_1|K_1 + |\lambda_2|K_2)$-semirandom.

**Remark.** We say that $f : X \to Y$ is a partial map if it is defined on some measurable subset $X_1 \subset X$. The notion of a semirandom maps generalizes to the case of partial maps by letting a partial map $f : X \to Y$ be $K$-semirandom if the restriction $f : X_1 \to Y$ is $K$-semirandom, where the corresponding measure on $X_1$ is the restriction of the measure from $X$. Every statement we make about semirandom maps has its version for a partial semirandom map. In particular, if $f : X \to Y$ is $K$-semirandom, then the restriction of $f$ onto any $X_1 \subset X$ is $K$-semirandom. Moreover, trivial partial maps (those that are defined on an empty set) are $K$-semirandom for any $K \geq 0$.

A measure class on a space $X$ is a subset of $\mathcal{M}(X)$ where $\mathcal{M}(X)$ is the set of measures on $X$.

**Definition 10.3.** Let $X$ and $Y$ be measure spaces, and let $\mathcal{M} \subset \mathcal{M}(X)$ and $\mathcal{N} \subset \mathcal{M}(Y)$ be measures classes on $X$ and $Y$ respectively. (All measures from $\mathcal{M}$ and $\mathcal{N}$ are positive measures.) We say $f : X \to Y$ is $K$ semirandom with respect to $\mathcal{M}$ and $\mathcal{N}$ if for every $\mu \in \mathcal{M}$ there is $\nu \in \mathcal{N}$ such that $f$ is $K$-semirandom with respect to $\mu$ and $\nu$; that is, $f_*\mu \leq K\nu$. 
In a similar fashion as above, we define the notion of a semirandom map $f : \mathbb{R}X \to \mathbb{R}Y$ with respect to classes of measures $\mathcal{M}$ and $\mathcal{N}$ on $X$ and $Y$ respectively. The following proposition follows from Proposition 10.1.

**Proposition 10.3.** Let $X$, $Y$ and $Z$ denote three measure spaces, with classes of measures $\mathcal{M}$, $\mathcal{N}$ and $\mathcal{Z}$ respectively. If $f : \mathbb{R}X \to \mathbb{R}Y$ is $K$-semirandom with respect to $\mathcal{M}$ and $\mathcal{N}$, and $f : \mathbb{R}Y \to \mathbb{R}Z$ is $L$-semirandom with respect to $\mathcal{N}$ and $\mathcal{Z}$, then $g \circ f : \mathbb{R}X \to \mathbb{R}Z$ is $KL$-semirandom with respect to $\mathcal{M}$ and $\mathcal{Z}$.

We say that a class of measures $\mathcal{M}$ is convex if it contains all convex combinations of its elements. The following proposition then follows from Proposition 10.2.

**Proposition 10.4.** If $f_i : \mathbb{R}X \to \mathbb{R}Y$ is $K_i$-semirandom with respect to classes of measures $\mathcal{M}$ and $\mathcal{N}$, $i = 1, 2$, and if $\mathcal{N}$ is convex, then for $\lambda_i \in \mathbb{R}$, the map $(\lambda_1 f_1 + \lambda_2 f_2) : \mathbb{R}X \to \mathbb{R}Y$ is $(|\lambda_1|K_1 + |\lambda_2|K_2)$-semirandom with respect to $\mathcal{M}$ and $\mathcal{N}$.

**Remark.** The space $\mathbb{R}X$ is naturally contained in the space $\mathcal{M}(X)$, and in a similar way we can define the notion of a semirandom map $f : \mathcal{M}(X) \to \mathcal{M}(Y)$.

**Natural measure classes.** Let $X_i$, $i = 1, \ldots, k$, denote measure spaces with measure classes $\mathcal{M}_i$. Let $X_1 \times \cdots \times X_k$ denote the product space, and by $\pi_i : (X_1 \times \cdots \times X_k) \to X_i$ denote the coordinate projections. By $\mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_k$, we denote the set of measures on $X_1 \times \cdots \times X_k$ that arise as the convex combinations of all standard products $\mu_1 \times \cdots \times \mu_k$ with $\mu_i \in \mathcal{M}_i$. We also define a natural class of measures $\mathcal{M}_1 \boxtimes \mathcal{M}_2 \boxtimes \cdots \boxtimes \mathcal{M}_k$ on $X_1 \times \cdots \times X_k$ as

\[
\mathcal{M}_1 \boxtimes \mathcal{M}_2 \boxtimes \cdots \boxtimes \mathcal{M}_k = \{ \mu \in \mathcal{M}(X_1 \times \cdots \times X_k) : (\forall i)(\exists \mu_i \in \mathcal{M}_i)((\pi_i)_* \mu = \mu_i) ) \}.
\]

This produces a large class of measures even if each $\mathcal{M}_i$ consists of a single measure. If each $\mathcal{M}_i$ is convex, then $\mathcal{M}_1 \boxtimes \mathcal{M}_2 \boxtimes \cdots \boxtimes \mathcal{M}_k$ is as well. If $X_1 = X$ and $M_i = \mathcal{M}$, then the standard product measure on $X^k$ is $\mathcal{M}^{\times k}$ and the other class of measures is denoted by $\mathcal{M}^{\boxtimes k}$.

We define the class $\mathcal{L}_1$ of Borel measures on $\mathbb{R}$ by saying that $\mu \in \mathcal{L}_1$ if $\mu(x, x+1) \leq 1$ for all $x \in \mathbb{R}$. This is a closed convex class of measures. Likewise we define the class of measures $\mathcal{L}_1$ on $\mathbb{R}/\lambda\mathbb{Z}$ for $\lambda > 1$ by saying that $\mu \in \mathcal{L}_1$ if $\mu(x, x+1) \leq 1$ for all $x \in \mathbb{R}/\lambda\mathbb{Z}$. The class of measures $\mathcal{L}_1$ is the class of measures that are controlled by the Lebesgue measure at the unit scale.

We consider the following spaces and their measure classes. In this paper, we define several maps (or partial maps) between these spaces (or their powers) and prove they are semirandom. We have
(i) The space of curves $\Gamma_{1,R}$ with the measure class containing the single measure $\sigma_T$ that is defined by setting $\sigma_T(\gamma) = Re^{-2R}$ for every $\gamma \in \Gamma_{1,R}$. We may assume that $\varepsilon$ is small enough so that $\Gamma_{\varepsilon,R} \subset \Gamma_{1,R}$.

(ii) The space of pants $\Pi_{1,R}$ with the measure class containing the single measure $\sigma_\Pi$ given by $\sigma_\Pi(\Pi) = e^{-3R}$. We may assume that $\varepsilon$ is small enough so that $\Pi_{300\varepsilon,R} \subset \Pi_{1,R}$.

(iii) Let $\Gamma_{1,R} = \{(x,\gamma) : \gamma \in \Gamma_{1,R}, x \in T_\gamma\}$ denote the space of pointed curves. (Recall that $T_\gamma = \mathbb{R}/I(\gamma)\mathbb{Z}$ is the parametrizing torus for $\gamma$.) The space $\hat{\Gamma}_{1,R}$ is really just the union of parametrizing tori $T_\gamma$ for curves $\gamma \in \Gamma_{1,R}$. By $\Sigma_\Gamma$ we denote the measure class on $\hat{\Gamma}_{1,R}$, such that $\mu \in \Sigma_\Gamma$ if the restriction $\mu_\gamma = \mu|_{T_\gamma}$ is in $e^{-2R}\mathcal{L}_1$, where $\mathcal{L}_1$ is the measure class on the circle $T_\gamma$ that was defined above.

(iv) Let $\Gamma_{1,R}^k = \{(x_1, \ldots, x_k, \gamma) : \gamma \in \Gamma_{1,R}, x_i \in T_\gamma\}$ denote the space of curves with $k$ marked points. The space $\hat{\Gamma}_{1,R}^k$ is canonically contained in $\left(\hat{\Gamma}_{1,R}\right)^k$.

The measure class $\Sigma_{\Gamma}^k$ on $\hat{\Gamma}_{1,R}^k$ is the restriction of $\Sigma_{\Gamma}^{2k}$ on the image of $\hat{\Gamma}_{1,R}$ in $\left(\hat{\Gamma}_{1,R}\right)^k$.

(v) The space $G = \pi_1(S,*)$ with the measure class $\Sigma_G$ that is the convex closure of the collection of measures $\sigma_a$ on $G$, where $\sigma_a$ is defined so that for $X \in G$, we have $\sigma_a(X) = \nu_a(I(\cdot X \cdot))e^{-I(\cdot X \cdot)}$, where $\nu_a(x) = 1$ if $x \in [a, a+1]$, and $\nu_a(x) = 0$ otherwise.

We observe that there exists a constant $K = K(S)$ such that for any measure $\mu$ in any of the above defined measure classes, the total measure of $\mu$ is bounded by $K$.

Finally we consider the map $\partial : \Pi_{1,R} \to \mathbb{R}\Gamma_{1,R}$ defined by $\partial \Pi = \gamma_0 + \gamma_1 + \gamma_2$, where $\gamma_i$ are the three oriented boundary curves of $\Pi$. We observe that $\partial$ is $K(S)$-semirandom from $\sigma_\Pi$ to $\sigma_T$.

*Standard maps are semirandom.* We consider several standard mappings and prove they are semirandom.

**Lemma 10.1.** Let $l > 0$ and $a, b \leq l - 1$. Then for any $Z \in G = \pi_1(S,*)$ such that $I(\cdot Z \cdot) = l$, there are at most $Ke^{\frac{a+b+1}{l}}$ ways of writing $Z = XY$, with $I(\cdot X \cdot) \in [a, a+1]$ and $I(\cdot Y \cdot) \in [b, b+1]$, for some $K = K(S)$.

**Proof.** Suppose that $X$ and $Y$ satisfy the given conditions. Consider a triangle in $\mathbb{H}^2$ whose sides are lifts of $\cdot X \cdot$, $\cdot Y \cdot$ and $\cdot Z \cdot$. (These lifts are denoted the same as the arcs we are lifting.) Then we drop the perpendicular $t$ from the vertex $z$ opposite to $\cdot Z \cdot$ to the side $\cdot Z \cdot$, and we let $a'$ and $b'$ be the lengths of the subintervals of $\cdot Z \cdot$ that meet at the endpoint of $t$ on $\cdot Z \cdot$.  

(then $a' + b' = I(Z \cdot)$). For simplicity, set $t = I(t)$. We find that

$$a \leq I(X \cdot) \leq t + a' \leq I(X \cdot) + \log 2 \leq a + 2$$

and, likewise, $b \leq t + b' \leq b + 2$. So

$$t \in \left[ \frac{a + b - l}{2}, \frac{a + b - l}{2} + 2 \right]$$

and

$$a' \in \left[ \frac{a - b + l}{2} - 2, \frac{a - b + l}{2} + 2 \right].$$

Therefore, the vertex $z$ must lie in a disc of radius $a + b - l + 4$ around the point on $Z$ that is $\frac{a - b + l}{2}$ away from the initial point of $Z$. It follows that there are at most $K(S)e^{\frac{a + b - l}{2}}$ lifts of the base point in this disc, and we are finished. \(\square\)

Let $p : G \times G \to G$ be the product map; that is, $g(X,Y) = XY$.

**Lemma 10.2.** The map $p : G \times G \to G$ is $K$-semirandom with respect to $\Sigma_G \times \Sigma_G$ on $G^2$ and $\Sigma_G$ on $G$ for some $K = K(S)$.

**Proof.** Let $a, b \in [0, \infty)$, and assume $b \geq a$. Recall the measures $\sigma_a$ on $G$, and let $\sigma = p_*(\sigma_a \times \sigma_b)$. We must show that $\sigma \leq K\Sigma_G$.

Let $Z \in G$, and let $l = I(Z \cdot)$. If $a \leq b \leq l - 1$, then

$$\sigma(Z) \leq K e^{\frac{a + b - l}{2}} e^{-a} e^{-b} = K e^{l} e^{-a - b - l}.$$ (If $l > a + b + 2$, then $\sigma(Z) = 0$.)

If $l - 1 \leq b$, then because there are at most $K e^{a} X$’s in $G$ for which $\sigma_a(X) > 0$, we find

$$\sigma(Z) \leq K e^{a} e^{-a} e^{-b} = K e^{-l} e^{-(b-l)}.$$ Then we see that

$$\frac{1}{K} \sigma \leq \sum_{k=\lfloor b-a-1 \rfloor}^{\lfloor b+1 \rfloor} e^{-(b-k)} \sigma_k + \sum_{k=\lceil b \rceil}^{\lfloor a+b+3 \rfloor} e^{-\frac{(a+b-k)}{2}} \sigma_k,$$ so $\sigma \leq K\Sigma_G$. \(\square\)

We define a partial map proj : $G \to \hat{\Gamma}_{1,R}$ as follows. Given $A \in G$, we let $\gamma = [A]$, and we let $z \in \gamma$ be the projection of the base point $* \in \gamma$. As always, the projection is defined by choosing lifts of $A \cdot \gamma$ in $H^2$ that have the same endpoints. Then we project a lift of $*$ to the lift of $\gamma$, where the lift of $* \in \gamma$ belongs to the lift of $\gamma$, and we let proj$(A) = (\gamma, z)$.

**Lemma 10.3.** The map $\text{proj} : G \to \hat{\Gamma}_{1,R}$ is $K(S)$-semirandom with respect to $\Sigma_G$ and $\Sigma_{\Gamma}$. 

Proof. Let $J$ be a unit interval on a curve $\gamma \in \Gamma_{1,R}$. We have seen in the two previous proofs that there are at most $Ke^{\frac{\cdot-2R}{2}}$ many $Z \in G$ for which $l(Z) \leq l$, and $\text{proj}(Z) \in J$. Therefore, if $\sigma \in \Sigma_G$, then

$$\text{proj}_{\sigma}(\gamma, J) \leq K \sum_{k=\lfloor 2R \rfloor}^{\infty} e^{\frac{k-2R}{2}}e^{-k} \leq Ke^{-2R},$$

and we are finished. $\square$

Another standard map we consider is the projection map $\Gamma_{1,L} \to \Gamma_{1,L}$ given by $(\gamma, x) \to \gamma$. This map is clearly 1-semirandom. Going in the opposite direction, we have the map $\gamma \to (\gamma, x)$ that assigns to $\gamma \in \Gamma_{1,R}$ a random point $x \in \gamma$. This map is really defined as a map $\mathcal{M}(\Gamma_{1,R}) \to \mathcal{M}(\Gamma_{1,R})$, and we observe that it is 1-semirandom as well.

Remark. We also observe that for $T \in G$, the map $\{1\} \to G$ defined by $1 \to T$ is $e^{\mu(T)}$-semirandom with respect to the unit measure on $\{1\}$ and $\Sigma_G$.

The principles of randomization. After almost every lemma or theorem we prove in Sections 4–9, we have added a “Randomization remark” that considers the functions we have implicitly defined, states their domain and range, and argues that the functions are semirandom with respect to a certain measure class. In the remarks we have followed the following principles:

1. When we write “a random element” (of a finite set $S$) that the reader was previously told to read as “an arbitrary element,” we now mean “the random element” of $\mathbb{R}S$, namely,

$$\frac{1}{|S|} \sum_{x \in S} x.$$

If $a \in \mathbb{R}S \subset \mathcal{M}(S)$ and $\mathcal{M}$ is a measure class on $S$, we say that $a$ is a $K$-semirandom element of $S$, with respect to $\mathcal{M}$, if there exists $\mu \in \mathcal{M}$ such that $a \leq K\mu$.

2. We can replace at will any map $f : X \to Y$ (or $f : X \to \mathbb{R}Y$) by the linear extension $f : \mathbb{R}X \to \mathbb{R}Y$. This can cause confusion if you think about it the wrong way, so we offer the following example to clarify what is going on.

In the hard case of the GSL, we take a random third connection (meaning the random third connection) and then cancel out one square $(A_{ij})$, $i, j = 0, 1$, of boundaries to get a formal sum of squares $(B_{ij})$ of curves. We then find for each new square (in the formal sum) a second third connection at random from a set depending on $(B_{ij})$ to complete the argument. The right way to think of the randomization (and linearization) is that the first operation defines a partial map $q_1 : (\Gamma_{1,R})^4 \to \mathbb{R}(\Gamma_{1,R})^4$. 
and the second operation defines
\[ g_0 : \Gamma_{1,R}^4 \to \mathbb{R}\Pi_{1,R}, \]
so we can write \( g_0 \circ q_1 \) by extending \( g_0 \) to a map from \( \mathbb{R}\Gamma_{1,R}^4 \) to \( \mathbb{R}\Pi_{1,R} \) linearly.

The danger is that one may try to imagine \( g_0 \) acting on a formal sum of curves by taking the random element from \( \mathbb{R}\text{Conn}_{\epsilon,R}(\cdot, \cdot) \).

So we will imagine that we are defining functions from \( X \) to \( Y \), or from \( X \) to \( \mathbb{R}Y \), and only think of them as functions from \( \mathbb{R}X \) to \( \mathbb{R}Y \) when we want to compose them.

3. We want to use the measure class \( \Sigma_G \times \Sigma_G = \Sigma^2_G \times \Sigma_G \) on \( G^2 = \{(X,Y)\} \) when we want to form the product \( XY \). We want to use the measure class \( \Sigma_G \boxtimes \Sigma_G = \Sigma^2_G \times \Sigma^2_G \) on \( G^2 \) if we want to be able to let \( X = Z \) and \( Y = Z \) for some \( Z \in G \).

For example, for the ASL, we use the measure class \( \Sigma^2_G \times \Sigma_G \times \Sigma^2_G \times \Sigma_G \) on six-tuples \((A_0, A_1, U, B_0, B_1, V)\) in \( G^6 = G^2 \times G \times G^2 \times G \). This is basically the largest measure class for which the maps \( \pi_{ij} : G^6 \to G^4 \) defined by \( \pi_{ij}(A_0, A_1, U, B_0, B_1, V) = (A_i, U, B_j, V) \) are 1-semirandom with respect to the measure class \( \Sigma^4_G \times \Sigma^4_G \) on \( G^4 \).

This is exactly what we want, because we have to form the words \( A_i U B_j V \), but we need the freedom to assign to \( A_0 \) and \( A_1 \) (or \( B_0 \) and \( B_1 \)) the same value.

11. Appendix 2

We develop the theory of equidistribution and counting, based on the uniformly exponential mixing of the geodesic flow, that we need to prove Theorem 3.2.

11.1. Left and right actions. The group \( \text{PSL}(2, \mathbb{R}) \) acts on the unit tangent bundle \( T^1 \mathbb{H}^2 \) on the left. (We refer to this action as the action by isometries.) Namely, if \( v \in T^1 \mathbb{H}^2 \) and \( h \in \text{PSL}(2, \mathbb{R}) \), then \( h \cdot v = h(v) \) is the resulting vector in \( T^1 \mathbb{H}^2 \). Moreover, if \( u \) and \( v \) are two vectors in \( T^1 \mathbb{H}^2 \), then there exists a unique element \( h \in \text{PSL}(2, \mathbb{R}) \) such that \( h \cdot v = u \). This enables us to identify the unit tangent bundle \( T^1 \mathbb{H}^2 \) with \( \text{PSL}(2, \mathbb{R}) \) as follows. Choose a vector \( v_0 \in T^1 \mathbb{H}^2 \). We identify \( v_0 \) with the identity element \( 1 \) in \( \text{PSL}(2, \mathbb{R}) \). A vector \( v \in T^1 \mathbb{H}^2 \) is identified with the unique element \( h \in \text{PSL}(2, \mathbb{R}) \) so that \( h \cdot v_0 = v \).

The existence and uniqueness of this \( h \) means that we can define a right action of \( \text{PSL}(2, \mathbb{R}) \) on \( T^1 \mathbb{H}^2 \) by the equation \((h \cdot v_0) \cdot g = hg \cdot v_0 \). This right action is the unique right action that satisfies the following two properties:

(i) \( h \cdot v_0 = v_0 \cdot h \);

(ii) \( g \cdot (v \cdot h) = (g \cdot v) \cdot h \)

for any \( h, g \in \text{PSL}(2, \mathbb{R}) \).
By an instruction we mean any transformation \( q : T^1 \mathbb{H}^2 \to T^1 \mathbb{H}^2 \) that is natural in the sense that
\[ q(g \cdot v) = g \cdot q(v) \]
for any \( g \in \text{PSL}(2, \mathbb{R}) \). For example, the time \( t \) geodesic flow \( g_t \) and counterclockwise rotation by angle \( t \) (which we denote by \( R_t \)) are instructions. It follows from the previous paragraph that the instructions are exactly the maps of the form
\[ q(v) = v \cdot h \]
for some \( h \in \text{PSL}(2, \mathbb{R}) \).

We define the instruction \( Y_t \) by
\[ Y_t = R_{-\pi/2} \cdot g_t \cdot R_{\pi/2}, \]
so \( Y_t \) is the “rightward normal flow,” one rotates to the right for 90 degrees, then flows for the time \( t \), and then rotates to the left for 90 degrees.

**Lemma 11.1.** Any \( h \in \text{PSL}(2, \mathbb{R}) \) can be uniquely written as
\[ h = Y_a \cdot g_t \cdot R_c, \]
for some choice of \( a, t, c \in \mathbb{R} \).

**Proof.** We let \( v = v_0 \cdot h \); we want to find \( a, b \) and \( c \) such that \( v = Y_a \cdot g_t \cdot R_c \).

Let \( p_0 \) and \( p \) be the base points of \( v_0 \) and \( v \). We let \( \alpha \) be the geodesic through \( p_0 \) orthogonal to \( v_0 \) (oriented to point to the right), and we let \( q \) be the closest point on \( \alpha \) to \( p \). Then, as shown in the Figure 11.1,

- we let \( a \) be the signed distance from \( p_0 \) to \( q \);
- we let \( \beta \) be the oriented geodesic through \( v_0 \cdot Y_a \), and let \( t \) be the signed distance along \( \beta \) from \( q \) to \( p \);
- we let \( c \) be the counter clockwise angle from \( v_0 \cdot Y_a \cdot g_t \) to \( v \).

Then we have \( v = v_0 \cdot Y_a \cdot g_t \cdot R_c \). \( \square \)

We equip \( \text{PSL}(2, \mathbb{R}) \) with the following distance function. Let \( h_j \in \text{PSL}(2, \mathbb{R}), \ j = 1, 2, \) and let \( v_j \in T^1 \mathbb{H}^2 \) denote the vectors corresponding to \( h_j \). (The vector \( v_j \) is based at the point \( p_j \in \mathbb{H}^2 \).) We let
\[ d_{\text{PSL}(2, \mathbb{R})}(h_1, h_2) = d(p_1, p_2) + \Theta(u_1, v_2), \]
where \( u_1 \in T^1 \mathbb{H}^2 \) is the parallel transport of the vector \( v_1 \) at the point \( p_2 \).

(Recall that \( \Theta(u, v) \) is the unoriented angle between vectors \( u \) and \( v \).) For \( h \in \text{PSL}(2, \mathbb{R}) \), we denote by \( ||h|| \) the distance between \( h \) and the identity element \( 1 \in \text{PSL}(2, \mathbb{R}) \).

We leave the proof of the following lemma to the reader.

**Lemma 11.2.** There are universal constants \( \delta_0, K_0 > 0 \) such that providing \( ||h|| \leq \delta_0 \), then
\[
h \cdot g_t = Y_a \cdot g_{t+b} \cdot R_c,
\]
where \( |a| + |b| + |c| \leq K_0 ||h|| \).

We now discuss the equidistribution of the equidistant lines on a closed Riemann surface \( S \). Let \( \alpha : \mathbb{R} \to S \) be a unit speed geodesic, and let \( \hat{\alpha} : \mathbb{R} \to T^1 S \) be the leftward normal unit vector field, given by \( \hat{\alpha}(s) = i\alpha'(s) \). (Here \( i \) denotes the imaginary unit in the tangent space to \( S \) at the point \( \alpha(s) \in S \).)

Let \( t \in \mathbb{R} \), and consider the vector field \( g_t(\hat{\alpha}) \). Then the vectors from the field \( g_t(\hat{\alpha}) \) are orthogonal to the line that is equidistant (at distance \( t \)) from the geodesic \( \alpha \).

We let \( dV \) be the volume form on \( T^1 S \), normalised so that
\[
\int_{T^1 S} dV = 1.
\]

The following theorem provides explicit estimates for how evenly distributed \( g_t(\hat{\alpha}) \) is in \( T^1 S \).

**Theorem 11.1.** Let \( f : T^1 S \to \mathbb{R} \) be any \( C^1 \) function. Then for \( a \geq C_1 e^{-qt} \), we have
\[
\left| \frac{1}{a} \int_0^a f(g_t(\hat{\alpha}(s))) \, ds - \int_{T^1 S} f \, dV \right| \leq C_2 e^{-\frac{qt}{a}} \left( \frac{1}{a} + ||f||_{C^1} \right),
\]
where the positive constants \( C_1, C_2 \) and \( q \) depend only on \( S \).

**Proof.** We let \( \psi_\eta : \text{PSL}(2, \mathbb{R}) \to [0, \infty) \) be such that
(i) \( \psi_\eta \) is supported in \( B_\eta(1) \), which is the ball of radius \( \eta \) centered at \( 1 \);
(ii) \( \int \psi_\eta = 1 \), where we integrate with respect to the Haar measure on \( \text{PSL}(2, \mathbb{R}) \);
(iii) \( ||\psi_\eta||_{C^1} \leq K_1 \eta^{-4} \) for some universal constant \( K_1 \);
(iv) \( \psi_\eta(X) = \psi_\eta(X^{-1}) \) for \( X \in \text{PSL}(2, \mathbb{R}) \).

We can arrange that (iii) holds because \( \psi_\eta \) needs to reach the height of \( \eta^{-3} \) in a space of size \( \eta \) (so the derivative of \( \psi_\eta \) is proportional to \( \eta^{-4} \)). For simplicity, we let \( \psi = \psi_\eta \).

If \( u, v \in T^1 \mathbb{H}^2 \), then there is a unique \( g \in \text{PSL}(2, \mathbb{R}) \) such that \( u \cdot g = v \). We let \( \psi(u, v) = \psi(g) \). (Condition (iv) above implies that \( \psi(u, v) = \psi(v, u) \).

Then for $a < b$ and $X \in T^1S$, we let
\[
\hat{\alpha}_{a,b}(X) = \int_a^b \psi(\hat{\alpha}(s), X) \, ds,
\]
and we let $\hat{\alpha}_a = \hat{\alpha}_{0,a}$.

Then
\[
||\hat{\alpha}_{a,b}||_{C^1} \leq (b - a)||\psi||_{C^1}
\]
and
\[
\int_{T^1S} \hat{\alpha}_{a,b} \, dV = b - a.
\]

Applying the factorization lemma above (Lemma 11.2), we find that
\[
\int_{T^1S} f(X)\hat{\alpha}_a(g_{-t}X) \, dV(X) = \int_0^a \int_{B_\eta(1)} f(\hat{\alpha}(s) \cdot h \cdot g_t)\psi(h) \, ds \, dV(h)
\]
\[
= \int_0^a \int_{B_\eta(1)} f(\hat{\alpha}(s) \cdot Y_a \cdot g_{t+b} \cdot R_c)\psi(h) \, ds \, dV(h)
\]
\[
= \int_0^a \int_{B_\eta(1)} f(\hat{\alpha}(s + a(t,h)) \cdot g_{t+b(t,h)} \cdot R_{c(t,h)})\psi(h) \, ds \, dV(h)
\]
\[
= \int_0^a \int_{B_\eta(1)} f(\hat{\alpha}(s + a(t,h))\psi(h) \, ds \, dV(h) + O(||f||_{C^1} \eta a),
\]
where the last equality follows from the upper bounds on $b(t,h)$ and $c(t,h)$ from Lemma 11.2. This yields the inequalities
\[
\int_{-K_0\eta}^{a-K_0\eta} f(g_t(\hat{\alpha}(s))) \, ds - K_0||f||_{C^1} \eta \leq \int_{T^1S} f(X)\hat{\alpha}_a(g_{-t}X) \, dV(X)
\]
\[
\leq \int_{-K_0\eta}^{a+K_0\eta} f(g_t(\hat{\alpha}(s))) \, ds + K_0||f||_{C^1} \eta.
\]

On the other hand, by exponential mixing,
\[
\left|\int_{T^1S} ((g_t)_*\hat{\alpha}_a)(X)f(X) \, dV(X) - \int_{T^1S} \hat{\alpha}_a \, dV \int_{T^1S} f \, dV \right|
\leq Ce^{-qt}||\hat{\alpha}_a||_{C^1}||f||_{C^1} \leq Ce^{-qt}a\eta^{-1}||f||_{C^1},
\]
where $C, q > 0$ depend only on $S$. So we obtain, for $a > 2K_0\eta$,

$$\int_0^a f(g_t(\alpha(s))) \, ds \leq \int_{T^1S} f((g_t)_* \hat{\alpha}_{-K_0\eta,a+K_0\eta}) \, dV + K_0 ||f||_{C^1} \eta$$

$$\leq (a + 2K_0\eta) \int_{T^1S} f \, dV + Ce^{-qt}a\eta^{-4}||f||_{C^1} + K_0\eta ||f||_{C^1}$$

and likewise

$$\int_0^a f(g_t(\alpha(s))) \, ds \geq (a - 2K_0\eta) \int_{T^1S} f \, dV - (Ce^{-qt}a\eta^{-4} + K_0\eta) ||f||_{C^1}.$$

Letting $\eta = e^{-\frac{1}{5}qt}$, the theorem follows. $\square$

11.2. Counting good connections. Let $\beta$ be another geodesic on $S$, and define $\hat{\beta} : \mathbb{R} \to T^1S$ in analogy to $\hat{\alpha}$. For intervals $I$ and $J$ in $\mathbb{R}$, we let

$$M_{I,J} = \{g_t(\hat{\beta}(s)) : (s, t) \in I \times J\}$$

be a 2-submanifold of $T^1S$.

For $J = [0, j_1]$, $M_{I,J}$ is the result of flowing out the normal field to $\hat{\beta}(s)$ for time $j_1$ along the geodesic flow.

We let $Y$ be the vector field associated to the rightward normal flow on $T^1S$ so that the time $t$ flow by $Y$ is equal to the instruction $Y_t$ on $T^1S$. Then $\hat{\alpha}'(s) = Y(\hat{\alpha}(s))$ for any geodesic $\alpha : \mathbb{R} \to T^1S$. We let

$$Y_t = \frac{1}{\cosh t} (g_t)_* Y,$$

so

$$\frac{\partial}{\partial s}(g_t(\hat{\alpha}(s))) = (\cosh t)Y_t(g_t(\hat{\alpha}(s))).$$

Then $Y^t$ is the unit speed rightward flow along normal fields to curves that are distance $t$ from the geodesic, where the normal field points away from the geodesic. For $t$ large, $Y^t$ is close to the negative horocyclic flow.

The following theorem provides a precise estimate on the weighted number of times that the equidistant curve $g_t(\alpha(s))$ (which is also a flow curve along $Y^t$) intersects the 2-manifold $M_{I,J}$.

**Theorem 11.2.** Let $f$ be a $C^1$ function with compact support on $M_{I,J}$. Then

$$\left| \frac{1}{a \cosh t} \sum_{s \in [0,a]} f(g_t(\hat{\alpha}(s))) - \int_{M_{I,J}} f \, \iota_{Y^t} \, dV \right|$$

$$\leq Ce^{-\frac{qt}{5}} \left( \frac{1}{a} + ||f||_{C^1} \right),$$
provided that $J = [j_0, j_1]$, where $|I|, |j_0|, |j_1| < \delta$, $t \geq 1$, and $1 \geq a \geq C_1 e^{-qt}$, where $C, C_1, \delta > 0$ depend only on $S$ and $\|f\|_{C^1} \leq 1$. (Here $i_Y dV$ is the contraction of the volume form $dV$ by the vector $Y^t$.)

Proof. The assumptions on $M_{I,J}, t$ and $a$ imply that the map $Q : M_{I,J} \times (0, \varepsilon) \to T^1 S$, defined by $Q(q, r, s) = Y_s (g_r (\tilde{\beta}(q)))$, is injective for some $\varepsilon = \varepsilon(S)$. We let $\psi$ be a $C^1$ bump function on $(0, \varepsilon)$, and we let $\tilde{f} : Q(M_{I,J} \times (0, \varepsilon)) \to \mathbb{R}$ be defined by

$$\tilde{f}(Q(q, r, s)) = f(g_r (\tilde{\beta}(q))) \psi(s).$$

Then

$$\|\tilde{f}\|_{C^1} \leq C(S) \|f\|_{C^1},$$

and

$$\int_{T^1 S} \tilde{f} dV = \int_M f i_Y dV.$$

Moreover,

$$\left| \sum_{g_r (\tilde{\alpha}(s)) \in M_{I,J}} f(g_t (\tilde{\alpha}(s))) - \cosh t \int_0^a \tilde{f}(g_t (\tilde{\alpha}(s))) ds \right| \leq \|f\|_{C^1}.$$

This inequality holds because every time the curve $Y_s (g_r (\tilde{\alpha}(s)))$ (for $s \in [0, \cosh t]$) crosses $M_{I,J}$, it goes through $Q$ (and contributes the same amount to the sum and the integral), except that the curve may start in $Q$ and miss $M_{I,J}$, and the terminal point may end in $Q$, contributing more to the sum than to the integral. For both endpoints, the error is at most $\|f\|_{C^1}$, and the error has different signs at the two endpoints, so the total error is at most $\|f\|_{C^1}$.

Therefore, by Theorem 11.1,

$$\left| \frac{1}{\cosh t} \sum_{g_r (\tilde{\alpha}(s)) \in M_{I,J}} f(g_t (\tilde{\alpha}(s))) - \int_{M_{I,J}} f i_Y dV \right| \leq \frac{1}{\cosh t} \|f\|_{C^1} + Ce^{-qt} \left( \frac{1}{a} + \|f\|_{C^1} \right) \leq Ce^{-qt} \left( \frac{1}{a} + \|f\|_{C^1} \right).$$

If $\alpha$ and $\beta$ are two geodesic segments, and $\varepsilon, L > 0$, we let $\text{Conn}_{\varepsilon,L}(\alpha, \beta)$ be the set of $(r, s, t)$ such that $g_t (\tilde{\alpha}(r)) = \tilde{\beta}(s)$ and $t \in [L, L + \varepsilon]$.

The following theorem, which is the object of this appendix, provides an estimate for the number of orthogeodesic connections, with length between $L$ and $L + \delta^2$, between two geodesic segments, where the length of the geodesic
segments, the interval length $\delta^2$, and the proportional error to the count are all exponentially small in $L$.

**Theorem 11.3.** Letting $\delta = e^{-\frac{qL}{40}}$, and $\alpha, \beta$ geodesic segments of length $\delta^2$, the number of orthogeodesics connections from one side of $\alpha$ to one side of $\beta$, of length in the interval $[L, L + \delta^2]$, is given by

$$\frac{1}{8\pi^2\chi(S)} \delta^6 e^L (1 + O(\delta)),$$

where the big $O$ constant depends only on $S$.

**Proof.** We let $M = M_{J^1}$, where $I = [0, \delta^2]$ and $J = [-\delta^2, 0]$. We want to count the number of $s \in [0, \delta^2]$ for which $g_L(\hat{\alpha}(s)) \in M$.

Let

$$M^+ = M_{[-\delta^3, \delta^2 + \delta^3], [-\delta^2 - \delta^3, \delta^3]}$$

be a slightly larger surface, and let $f^+$ be a $C^1$ function on $M^+$ that is equal to 1 on $M$. We can arrange

$$||f^+||_{C^1} \leq 10\delta^{-3},$$

and $f^+$ takes values in $[0, 1]$.

Then

$$\int_{M^+} |f^+ - \chi_M| \iota_{Y^L} dV \leq 10\delta^5,$$

and given our normalization of the Liouville volume form $dV$, we have

$$\int_M \iota_{Y^L} dV = \frac{1}{4\pi^2|\chi(S)|} \delta^4 + O(\delta^8).$$

Putting all this together and applying Theorem 11.2, we have

$$\frac{1}{\delta^2 \cosh L} \#\text{Conn}_{\epsilon,L}(\alpha, \beta) \leq \frac{1}{\delta^2 \cosh L} \sum_{g_L(\hat{\alpha}(s)) \in M^+} f^+(g_L(\hat{\alpha}(s)))$$

$$\leq \int_{M^+} f^+ \iota_{Y^L} dV + C\delta^8 \left(\frac{1}{\alpha} + ||f||_{C^1}\right)$$

$$\leq \frac{\delta^4}{4\pi^2|\chi(S)|} + O(\delta^8) + 10\delta^5 + C\delta^8(\delta^{-2} + \delta^{-3})$$

$$\leq \frac{\delta^4}{4\pi^2|\chi(S)|} + C\delta^5.$$
and prove that
\[ \frac{1}{\delta^2 \cosh L} \#\text{Conn}_{\varepsilon,L}(\alpha,\beta) \geq \frac{\delta^4}{4\pi^2|\chi(S)|} - C\delta^5. \]
Since \( \cosh L = \frac{e^L}{2} (1 + O(e^{-2L})) \), the theorem follows. \( \square \)

References


(Received: January 11, 2013) 
(Revised: July 7, 2014)