SHORT CLOSED GEODESICS ON CUSPED HYPERBOLIC SURFACES

THI HANH VO

Abstract. This article deals with the set of closed geodesics on complete finite type hyperbolic surfaces. For any non-negative integer $k$, we consider the set of closed geodesics that self-intersect at least $k$ times, and investigate those of minimal length. The main result is that, if the surface has at least one cusp, their self-intersection numbers are exactly $k$ for large enough $k$.

1. Introduction

Let $X$ be a complete hyperbolic surface of finite type. For a non-negative integer $k$, we focus on the set of closed geodesics on $X$ with at least $k$ self-intersection points. By discreteness of the length spectrum, there exists a $\gamma_k$ which is shortest among these geodesics. Our goal is to study the actual value of the self-intersection number of $\gamma_k$.

For $k = 0$, the set under consideration is the set of all closed geodesics on $X$, so this amounts to asking how many times does the shortest closed geodesic (the systole) of $X$ self-intersect. Unless $X$ is a thrice punctured sphere, the answer is 0. For $k = 1$, a result by Buser [6] states that the shortest non-simple closed geodesic on $X$ has one self-intersection point (it is a so-called figure eight geodesic). For $k \geq 2$, there are no exact values known.

Let $I(k, X)$ denote the maximum number of self-intersections of shortest closed geodesics on $X$ with at least $k$ self-intersections. By definition, $I(k, X) \geq k$. Basmajian [2] provided the first upper bounds on these numbers. He showed that they are bounded from above by constants depending on $k$ and the genus of $X$. The dependence on the topology is used to bound the lengths of curves in a pants decomposition via a theorem by Bers [5].

Erlandsson and Parlier [9] were able to get rid of the topological data. They proved that for all $k \geq 2$, $I(k, X)$ is bounded from above by a universal linear function in $k$. In fact, this result holds true for general surfaces (those with non-abelian fundamental group, including infinite area or infinite type surfaces). In the presence of cusps, they provided a sharper bound on $I(k, X)$, which this time depends on the geometry of $X$. This bound implies that the numbers of self-intersections behave asymptotically like $k$ for growing $k$ when $X$ has at least one cusp.

To the best of the author’s knowledge, there is not a single surface $X$ for which the equality $I(k, X) = k$ is known to hold for all $k$. Conversely, there are no surfaces for which the equality is known to fail for at least one $k \geq 1$. We show that if $X$ has a cusp, the equality holds for all large enough $k$. 

Date: December 23, 2019.

2010 Mathematics Subject Classification. Primary: 30F10. Secondary: 53C22.

Key words and phrases. closed geodesics, hyperbolic surfaces.
Theorem 1.1. Let $X$ be a complete finite type hyperbolic surface with at least one cusp. There exists a constant $K(X)$ depending on $X$ such that for all $k \geq K(X)$, if $\gamma_k$ is a shortest closed geodesic with self-intersection number at least $k$, then

(1) $\gamma_k$ self-intersects exactly $k$ times,
(2) $\gamma_k$ is freely homotopic to $\kappa * \tau^k$, where $\tau$ is freely homotopic to a cusp, and $\kappa$ is a non-contractible simple loop that is not freely homotopic to $\tau$.

The constant $K(X)$, which can be found explicitly in Section 3, comes from the systole length and the shortest orthogonal distances from horocycles to themselves. In our proof, we show that in terms of length, it is more efficient to wind around a cusp to generate self-intersection points than by doing anything else.

To simplify, we will use the term $ba^k c$ geodesic to refer to the geodesic mentioned in Theorem 1.1(2). Let $\mathcal{M}_{g,n}$ denote the set of all cusped hyperbolic surfaces of genus $g$ with $n$ punctures or boundary components, that are not the thrice punctured sphere, and the systole lengths of which are at least $\delta$, where $\delta > 0$. For $S \in \mathcal{M}_{g,n}$, by Theorem 1.1 there is a constant $K(S)$ depending on the systole length and the shortest orthogonal distances from horocycles to themselves, such that for all $k \geq K(S)$, we have the conclusions (1) and (2). The systole length is bounded from below by $\delta$, whereas the shortest orthogonal distances can be evaluated explicitly in terms of topological data via Bers’ constant \cite{Bers}. Therefore, $K(S)$ can be expressed as a constant depending only on $\delta$ and the topological data.

Corollary 1.2. For all $S \in \mathcal{M}_{g,n}$, there exists a constant $K = K(g, n, \delta)$ such that for all $k \geq K$, if $\gamma_k$ is a shortest closed geodesic on $S$ with at least $k$ self-intersection points, then

(1) $\gamma_k$ self-intersects exactly $k$ times,
(2) $\gamma_k$ is a $ba^k c$ geodesic.

The article is organized as follows. In Section 2 we generalize some computations on lengths and self-intersection numbers of geodesic segments. Section 3 is dedicated to the proof of Theorem 1.1. In Section 4 we prove Corollary 1.2 and present some remarks.

2. Preliminary computations

Throughout this paper, $X$ will be a complete finite type hyperbolic surface with at least one cusp. By abuse of language, we use the term curve to mean both a parameterized curve and the set of points that make up the image of a parameterized curve. Recall that a geodesic realizes the minimal number of self-intersections among all curves in its free homotopy class. All our closed geodesics are considered primitive.

Denote by $\mathcal{G}_k(X)$ the set of closed geodesics on $X$ that self-intersect exactly $k$ times. Basmajian \cite{Basmajian1, Basmajian2} investigated the following quantity

$$s_k(X) := \inf\{\ell(\gamma) : \gamma \in \mathcal{G}_k(X)\}.$$ 

He showed \cite{Basmajian1} Proposition 4.2 that for $k \geq 2$, there exists a constant

$$C(k, X) := 2\text{arsinh}(k) + d(X) + 1,$$

such that

$$s_k(X) \leq C(k, X),$$

(2.1)
where \( d(X) \) is the shortest orthogonal distance from the length one horocycle boundary of a cusp to the horocycle itself. Note that \( \text{arcsinh}(k) \) is comparable to \( \log(k) \), and hence so is \( C(k, X) \).

Likewise, denote by \( \mathcal{G}_{\geq k}(X) \) the set of closed geodesics on \( X \) that self-intersect at least \( k \) times. Consider the following quantity

\[
s_{\geq k}(X) := \inf \{ \ell(\gamma) : \gamma \in \mathcal{G}_{\geq k}(X) \}.
\]

Since \( s_{\geq k}(X) \leq s_k(X) \), inequality (2.1) also holds for \( s_{\geq k}(X) \). Thus, if \( \gamma_k \) is a shortest geodesic in \( \mathcal{G}_{\geq k}(X) \), then

\[
\ell(\gamma_k) \leq C(k, X).
\] (2.2)

Given \( \varepsilon > 0 \), we recall that the \( \varepsilon \)-thick part \( X_{\varepsilon T} \) of \( X \) is the subset of \( X \) consisting of points with injectivity radius at least \( \varepsilon \), and the \( \varepsilon \)-thin part \( X_{\varepsilon t} \) of \( X \) is the set of points with injectivity radius less than \( \varepsilon \). A curve \( \gamma \subset X \) can be decomposed into \( \gamma_{\varepsilon T} := X_{\varepsilon T} \cap \gamma \) and \( \gamma_{\varepsilon t} := X_{\varepsilon t} \cap \gamma \). Erlandsson and Parlier managed to control the relationship between length and self-intersection of \( \gamma_{\varepsilon T} \). They showed \([9, \text{Theorem 1.3}]\) that for \( \varepsilon \leq \frac{1}{2} \), the self-intersection of \( \gamma_{\varepsilon T} \) satisfies

\[
\ell(\gamma_{\varepsilon T}) > \frac{\varepsilon}{12} \sqrt{i(\gamma_{\varepsilon T}, \gamma_{\varepsilon T})}.
\] (2.3)

Now, let \( \varepsilon := \min\{\frac{1}{2}, \frac{\delta}{2}\} \), where \( \delta = \delta(X) \) is the systole length of \( X \). Choose \( K(X) \geq 2 \) such that

\[
C(k, X) < \frac{\varepsilon}{12} \sqrt{k},
\]

for all \( k \geq K(X) \). Let \( k \geq K(X) \) and let \( \gamma_k \) be a shortest geodesic in \( \mathcal{G}_{\geq k}(X) \). By inequalities (2.2) and (2.3), \( \gamma_k \) must enter \( X_{\varepsilon t} \), the \( \varepsilon \)-thin part of \( X \). Indeed, if \( \gamma_k \) is entirely contained in \( X_{\varepsilon T} \), then \( \gamma_k = \gamma_{\varepsilon T} \). We would have

\[
\frac{\varepsilon}{12} \sqrt{k} \leq \frac{\varepsilon}{12} \sqrt{i(\gamma_k, \gamma_k)} \leq \ell(\gamma_k) \leq C(k, X) < \frac{\varepsilon}{12} \sqrt{k},
\]

which is a contradiction. Thus, \( \gamma_k \) must enter \( X_{\varepsilon t} \). Due to our choice of \( \varepsilon \), \( \gamma_k \) must enter a cusp of \( X \). Hence, understanding the behavior of geodesic segments lying in cusp neighborhoods will be crucial. In the following, we present some facts about these geodesic segments.

For every cusp \( \mathcal{C} \) of \( X \), we have cusp neighborhoods filled by embedded horocycles winding around the cusp. Let \( h = h(\mathcal{C}) \) be the supremum of the lengths of embedded horocycles in \( X \) winding around \( \mathcal{C} \). Given \( 0 < h < h \), we denote by \( \mathcal{B}_h \) the cusp neighborhood with boundary given by an embedded horocycle of length \( h \). Let \( \beta \subset \mathcal{B}_h \) be a geodesic segment that starts on \( \partial \mathcal{B}_h \), winds around the cusp, and returns to \( \partial \mathcal{B}_h \). We refer to this type of geodesic segments as a strand (see Figure 1).

We define the winding number \( \omega(\beta) \) of the strand \( \beta \) in \( \mathcal{B}_h \) (with respect to \( \mathcal{B}_h \)) in the following way. Every point of \( \beta \) projects orthogonally to a well-defined point of the length \( h \) horocycle. The winding number of \( \beta \) is given by the ceiling of the quotient of the length of the projection of \( \beta \) (thought of as a parameterized segment) divided by \( h \).

In the sequel, the symbol \([\cdot]\) denotes the floor function. We begin with the following lemma which is a generalization of \([3, \text{Lemma 4.1}]\).
Lemma 2.1. For $0 < h \leq h^*$, let $\beta \subset \mathcal{B}_h$ be a strand that starts on $\partial \mathcal{B}_h$, winds $\omega(\beta)$ times around the cusp, and returns to $\partial \mathcal{B}_h$. Then

$$\omega(\beta) = i(\beta, \beta) + 1 = \left\lfloor \frac{2}{h} \sinh \frac{\ell(\beta)}{2} \right\rfloor.$$ 

Proof. The procedure is essentially the same as the proof of [3 Lemma 4.1]. Upon lifting to the upper half-plane model for the hyperbolic plane, we can normalize so that $\mathcal{B}_h$ lifts to the region above the Euclidean line $y = \frac{1}{h}$, and the parabolic element identified with the element of the fundamental group that wraps once around $\mathcal{B}_h$ is $f : z \mapsto z + 1$. The lift of $\beta$ in the upper half-plane is a geodesic segment $\tilde{\beta}$ that is contained on an Euclidean semicircle orthogonal to the real axis. This semicircle can be further normalized so that its center is the origin (see Figure 2). Denote the endpoints of this semicircle by $-r$ and $r$. Write the endpoints of $\tilde{\beta}$ as $-re^{i\theta_0}$ and $re^{-i\theta_0}$, where $\theta_0$ is given by

$$\cos(\theta_0) = \frac{\sqrt{r^2 - \frac{1}{h^2}}}{r}.$$ 

Parametrizing the geodesic segment joining $re^{i\theta_0}$ and $ri$ by $\tilde{\beta}(t) = re^{it}$ for $\theta_0 \leq t \leq \frac{\pi}{2}$, a computation yields

$$r = \frac{1}{h} \cosh \frac{\ell(\beta)}{2}. \quad (2.4)$$

Since $\beta$ winds around the cusp $\omega(\beta)$ times, the $f$-translates of $\tilde{\beta}$ intersect $\tilde{\beta}$ exactly $(\omega(\beta) - 1)$ times. Hence,

$$\omega(\beta) - 1 = i(\beta, \beta).$$
Thus,
\[
\omega(\beta) = \max \left\{ n \in \mathbb{N} : -r + n \leq -r + 2\sqrt{r^2 - \frac{1}{h^2}} \right\} = \left\lfloor 2\sqrt{r^2 - \frac{1}{h^2}} \right\rfloor.
\]  
(2.5)
Plugging \( r \) from equation (2.4) into the expression (2.5) allows to conclude after simplification.

As a direct consequence of the above lemma, we have the following corollary.

**Corollary 2.2.** For \( 0 < h < \tilde{h} \), let \( \beta \subset \mathcal{B}_h \) be a strand that starts on \( \partial \mathcal{B}_h \), winds \( \omega(\beta) \) times around the cusp, and returns to \( \partial \mathcal{B}_h \). Then
\[
2 \arcsinh \frac{h(\omega(\beta) - 1)}{2} \leq \ell(\beta) \leq 2 \arcsinh \frac{h\omega(\beta)}{2}.
\]

Using the same computation method as in the previous results, we have the following estimate.

**Lemma 2.3.** For \( 0 < h_0 < h < \tilde{h} \), let \( \beta \subset \mathcal{B}_h \) be a strand that has its endpoints on the horocycle of length \( h \). If the length of \( \beta \) satisfies
\[
\ell(\beta) > 2 \text{arccosh} \frac{h}{h_0},
\]
then \( \beta \) must enter \( \mathcal{B}_{h_0} \).

**Proof.** The strand \( \beta \) enters \( \mathcal{B}_{h_0} \) if and only if it is longer than \( \beta_0 \subset \mathcal{B}_h \), a strand that has its endpoints on \( \partial \mathcal{B}_h \), and meets the horocycle of length \( h_0 \) at the point of tangency. By the same computation method as in the proof of Lemma 2.1, the length of \( \beta_0 \) is given by
\[
\ell(\beta_0) = 2 \text{arccosh} \frac{h}{h_0},
\]
and hence the lemma follows.

In the following lemma, we estimate the difference between lengths of two strands that have the same winding number but stay in different levels of a cusp neighborhood.

**Lemma 2.4.** Let \( \beta_1 \subset \mathcal{B}_1 \) be a strand with endpoints on the horocycle of length \( 1 \) and for \( 0 < h < \frac{1}{10} \), let \( \beta_h \subset \mathcal{B}_h \) be a strand with endpoints on the horocycle of length \( h \). Suppose that both of them wind around the cusp \( \omega \) times. If \( \omega \geq \frac{1}{h} \), then
\[
\ell(\beta_1) > \ell(\beta_h) + 2 \arcsinh \frac{1}{6}.
\]

**Proof.** By Corollary 2.2, the lengths of \( \beta_1 \) and \( \beta_h \) satisfy
\[
\ell(\beta_1) \geq 2 \arcsinh \frac{\omega - 1}{2},
\]
(2.6)
and
\[
\ell(\beta_h) \leq 2 \arcsinh \frac{h\omega}{2}.
\]
(2.7)
Define the function
\[
f(x) := \arcsinh \frac{x - 1}{2} - \arcsinh \frac{hx}{2}.
\]
Since $f$ is increasing for $x \geq \frac{1}{h}$, we have
\[ f(\omega) \geq f\left(\frac{1}{h}\right). \]
Hence,
\[ \arcsinh \frac{\omega - 1}{2} - \arcsinh \frac{h\omega}{2} \geq \arcsinh \frac{\frac{1}{h} - 1}{2} - \arcsinh \frac{1}{2} \]
\[ \geq \arcsinh \frac{\frac{1}{h} - 1}{6} + \arcsinh(1) - \arcsinh \frac{1}{2} \quad (2.8) \]
\[ > \arcsinh \frac{1}{h} - \frac{1}{6}, \quad (2.9) \]
where inequality (2.8) holds since
\[ \arcsinh(u) + \arcsinh(v) < \arcsinh(3uv) \quad \text{for} \quad u, v \geq 1. \]
A combination of inequalities (2.6), (2.7) and (2.9) gives the desired result. □

We now consider sets of strands lying in the same cusp neighborhood. If $\beta_1, \beta_2 \subset \mathcal{B}_h$ are strands with endpoints on the horocycle of length $h$ and with winding numbers $\omega_1 \leq \omega_2$, it was shown in [9, Lemma 3.2] that $i(\beta_1, \beta_2) \leq 2\omega_1$. The following lemma is an extended version of this result.

Lemma 2.5. For $h < \tilde{h}$, let $\mathcal{S} \subset \mathcal{B}_h$ be a set consisting of $n \geq 2$ strands $\beta_1, \ldots, \beta_n$ with endpoints on the horocycle of length $h$ and winding numbers $\omega_1 \leq \cdots \leq \omega_n$, then
\[ i(\mathcal{S}, \mathcal{S}) \leq (2n - 1)\omega_1 + \cdots + 3\omega_{n-1} + \omega_n - n. \]

Proof. Using the fact that $i(\beta_i, \beta_i) = \omega_i - 1$ for all $i$ and $i(\beta_i, \beta_j) \leq 2\omega_i$ for all $i < j$, we have
\[ i(\mathcal{S}, \mathcal{S}) = i(\beta_1, \beta_1) + \cdots + i(\beta_n, \beta_n) + i(\beta_1, \beta_2) + \cdots + i(\beta_1, \beta_n) + \cdots + i(\beta_{n-1}, \beta_n) \]
\[ \leq \omega_1 + \cdots + \omega_n - n + 2(n - 1)\omega_1 + \cdots + 2\omega_{n-1}. \]
Simplifying finishes the proof. □

3. The proof of Theorem 1.1

We first define all the constants depending on $X$ that we will need in the proof. Let $\tilde{h} = \tilde{h}(X)$ be the length of the longest embedded horocycle boundary of a cusp of $X$, and $s = s(X)$ be the systole length of $X$. Set
\[ \tilde{h}_0 = \tilde{h}_0(X) := \frac{1}{30\tilde{h}}, \]
and
\[ \varepsilon_0 = \varepsilon_0(X) := \min\left\{ \frac{\tilde{h}_0}{5}, \frac{s}{5} \right\}. \]
Denote by $X_{\varepsilon_0T}$ the $\varepsilon_0$-thick part of $X$, and $X_{\varepsilon_0t}$ the $\varepsilon_0$-thin part of $X$. We can find the shortest orthogonal distance from each boundary component of $X_{\varepsilon_0t}$ to itself. Let $d(X)$ be the maximum value of such distances for all boundary components of $X_{\varepsilon_0t}$. Choose $D(X) \geq 2$ such that
\[ \frac{\varepsilon_0}{12} \sqrt{x} > 2 \arcsinh \frac{\varepsilon_0x}{2} + d(X) + 2\varepsilon_0, \quad \text{for all} \quad x \geq D(X). \quad (3.1) \]
Set
\[ \varepsilon = \varepsilon(X) := \frac{\varepsilon_0}{10D(X)}. \]
Denote by $X_{\varepsilon t}$ the $\varepsilon$-thick part of $X$, and $X_{\varepsilon t}$ the $\varepsilon$-thin part of $X$. Let $d(X)$ be the shortest orthogonal distance from the length one horocycle boundary of a cusp to the horocycle itself. Choose $K(X) \geq D(X)$ such that

$$C(k, X) := 2 \arcsinh(k) + d(X) + 1 < \frac{\varepsilon}{\rho} \sqrt{k}, \quad \text{for all } k \geq K(X).$$  \hfill (3.2)

Fix once and for all the constants $\varepsilon, \varepsilon_0, \rho_0, D(X)$ and $K(X)$. For all $k \geq K(X)$, we consider $\mathcal{G}_{\geq k}(X)$, the set of closed geodesics on $X$ that self-intersect at least $k$ times, and investigate those of minimal length. The following lemma is a key step in our proof.

**Lemma 3.1.** For all $k \geq K(X)$ and $\gamma_k$ a shortest geodesic in $\mathcal{G}_{\geq k}(X)$, $\gamma_k$ enters $X_{\varepsilon t}$ exactly once.

**Proof.** We first note that by inequalities (2.2) and (2.3), $\gamma_k$ must enter $X_{\varepsilon t}$. Since $\varepsilon < \varepsilon_0$, $\gamma_k$ must reach $X_{\varepsilon t}$ before entering $X_t$. Due to our choice of $\varepsilon_0$, $\gamma_k$ must enter a cusp neighborhood with boundary given by a length one horocycle (such a neighborhood is always embedded). Assume that $\gamma_k$ enters $m \geq 1$ such cusp neighborhood(s), say $\mathcal{B}_1, \ldots, \mathcal{B}_m$. For $1 \leq i \leq m$, let $\partial \mathcal{B}_i := \gamma_k \cap \partial \mathcal{B}_i$. Then $\partial \mathcal{B}_i$ is a set consisting of strands starting on $\partial \mathcal{B}_i$, winding around the respective cusp, and returning to $\partial \mathcal{B}_i$. Denote by $\gamma_i^1, \ldots, \gamma_i^{n_i}$ the strands in $\partial \mathcal{B}_i$ with winding numbers $\omega_i^1 \leq \cdots \leq \omega_i^{n_i}$. Let $\partial \mathcal{B}_i := \{ \delta_1, \ldots, \delta_n \}$, and $n := n_1 + \cdots + n_m \geq 1$. Recall that at least one strand of $\partial \mathcal{B}_i$ has to wind multiple times around a cusp to enter $X_{\varepsilon t} \subset X_{\varepsilon t}$. By our setting, it must cross a horocycle of length $\rho_0$. In what follows, we will show that there is only one such strand.

If $n = 1$, then the lemma holds because the strand of $\partial \mathcal{B}_i$ has to enter $X_{\varepsilon t}$. If $n \geq 2$, suppose that $\gamma_i^{n_1}$ and $\gamma_i^*$ are the two strands in $\partial \mathcal{B}_i$ which have the biggest winding numbers, $\omega_i^{n_1}$ and $\omega_i^*$, respectively. Assume that $\omega_i^{n_1} \geq \omega_i^*$. If $\omega_i^* < 17\rho_0$, we claim that $\gamma_i^*$ cannot enter the cusp neighborhood $\mathcal{B}_{\rho_0}$ with boundary a horocycle of length $\rho_0$. Indeed, if $\gamma_i^*$ enters the cusp neighborhood with boundary a horocycle of length $\rho_0$, then its length must be at least two times the distance between the horocycles of length $1$ and $\rho_0$, which is

$$2 \log \frac{1}{\rho_0} = 2 \log(30\rho_0),$$

whereas by Corollary 2.2,

$$\ell(\gamma_i^*) \leq 2 \arcsinh \frac{\omega_i^*}{2} < 2 \arcsinh \frac{17\rho_0}{2} < 2 \log(30\rho_0),$$

which is absurd. Thus, among all strand of $\partial \mathcal{B}_i$, only $\gamma_i^{n_1}$ enters a cusp neighborhood with boundary a horocycle of length $\rho_0$, and hence the lemma holds.

We now consider the case $\omega_i^* \geq 17\rho_0$. We first note that due to Lemma 2.5, an immediate upper bound on the self-intersection number of $\partial \mathcal{B}_i$ is given by

$$i(\partial \mathcal{B}_i, \partial \mathcal{B}_i) \leq (2n_1 - 1)\omega_1^1 + \cdots + \omega_1^{n_1} - n_1 + \cdots + (2n_m - 1)\omega_m^1 + \cdots + \omega_m^{n_m} - n_m.$$  \hfill (3.3)

We then proceed as follows.

- Evaluating the length of $\partial \mathcal{B}_i$ by constructing a new set of strands from it.
- Bounding the self-intersection number of $\partial \mathcal{B}_i$ and finalizing the proof.
Constructing a new set of strands from $\mathcal{S}$. The construction is an involved cut and paste type argument: by choosing a certain height level on the cusp regions, except for the longest strand $\gamma_1$, we cut all parts of the others which are above this level, and then paste them on top of $\gamma_1$, with the same orientation as the “winding” direction of $\gamma_1$. More precisely, let $\mathcal{S}'$ be a set consisting of $n$ strands, say $\gamma_i'$ for $1 \leq i \leq m$ and $1 \leq j \leq n_i$, such that each $\gamma_i'$ has the same initial and terminal points as $\gamma_i$, and has winding number $\omega_i'$ that is given by

$$
\omega_i' := \begin{cases}
(2n_1 - 1)\omega_1 + \cdots + \omega_{i-1}' + \cdots + (2n_m - 1)\omega_m + \cdots + \omega_{m-1}' + 1 & \text{if } (i, j) = (1, n_1), \\
\omega_i' & \text{if } \omega_i' < [3k], \\
\lfloor \omega_i' \rfloor & \text{otherwise},
\end{cases}
$$

where the symbol $\lfloor \cdot \rfloor$ denotes the ceiling function. Let $\gamma_k'$ be the geodesic in the free homotopy class of the piecewise geodesic closed curve $(\gamma_k \setminus \mathcal{S}) \cup \mathcal{S}'$. We claim that $\gamma_k' \in \mathcal{S}_{\geq k}(X)$. Indeed, for $\omega_i' \geq [3k], (i, j) \neq (1, n_1)$, by Corollary 2.2, the new strand $\gamma_i'$ satisfies

$$
\ell(\gamma_i') \geq 2 \arcsinh \frac{\omega_i' - 1}{2} \geq 2 \arcsinh \frac{3k - 1}{2} > 2 \arccosh(k).
$$

By Lemma 2.3, with this amount of length, even if it has to stand on the longest embedded horocycle around the cusp, it will still enter $B$. Concerning the other strands of $\mathcal{S}'$, they are either longer than the original strands (for $(i, j) = (1, n_1)$) or of the same length (otherwise). This fact guarantees that

$$
i(\gamma_k', \gamma_k') \geq i(\gamma_k, \gamma_k) \geq k.
$$

Since $\gamma_k$ realizes the minimal length in $\mathcal{S}_{\geq k}(X)$, we have $\ell(\gamma_k) \leq \ell(\gamma_k')$. As

$$
\ell(\gamma_k) = \ell(\gamma_k \setminus \mathcal{S}) + \ell(\mathcal{S}),
$$

and

$$
\ell(\gamma_k') \leq \ell(\gamma_k \setminus \mathcal{S} \cup \mathcal{S}') = \ell(\gamma_k \setminus \mathcal{S}) + \ell(\mathcal{S}'),
$$

we have

$$
\ell(\mathcal{S}) \leq \ell(\mathcal{S}').
$$

(3.4)

The length of $\mathcal{S}$ and $\mathcal{S}'$ are as follows:

$$
\ell(\mathcal{S}) = \ell(\mathcal{S}_1) + \cdots + \ell(\mathcal{S}_m)
$$

$$
= \ell(\gamma_1) + \cdots + \ell(\gamma_1') + \cdots + \ell(\gamma_m) + \cdots + \ell(\gamma_m'),
$$

and

$$
\ell(\mathcal{S}') = \ell(\mathcal{S}_1') + \cdots + \ell(\mathcal{S}_m')
$$

$$
= \ell(\gamma_1') + \cdots + \ell(\gamma_1') + \cdots + \ell(\gamma_m') + \cdots + \ell(\gamma_m').
$$

By our construction, the lengths of each pair of strands $\gamma_i$ and $\gamma_i'$ satisfy

$$
\ell(\gamma_i') = \ell(\gamma_i'), \quad \text{for } \omega_i' < [3k],
$$

(3.5)

and

$$
\ell(\gamma_i') \geq \ell(\gamma_i'), \quad \text{for } \omega_i' \geq [3k], (i, j) \neq \{(1, n), (*, *)\}.
$$

(3.6)

For $(i, j) = (*, *)$, recall that $\gamma_*$ is the strand in $\mathcal{S}$ that has the second biggest winding number $\omega_*$. Also recall that $\gamma_*'$ is a strand in $\mathcal{S}'$ that has the same endpoints as $\gamma_*'$ and
winds around the cusp \([3\hbar]\) times. Now, let \(\gamma^{n''}_s\) be a strand that has its endpoints on the horocycle of length \(\frac{1}{15\hbar}\) and that winds \(\omega^{n''}_s := \omega_s^*\) times around the cusp (see Figure 3).

![Figure 3. Strands \(\gamma^*_s, \gamma^{*'}_s\) and \(\gamma^{**}_s\).](image)

By Corollary 2.2 and Lemma 2.4 we have
\[
\ell(\gamma^*_s) \geq \ell(\gamma^{n''}_s) + 2\arcsinh \frac{15\hbar - 1}{6} \geq \ell(\gamma^{n''}_s) + 2\arcsinh \frac{3\hbar + 1}{2} \geq \ell(\gamma^{n''}_s) + \ell(\gamma^*_s), \tag{3.7}
\]
From inequalities (3.5), (3.6) and (3.7), we have
\[
\sum_{1 \leq i \leq m, 1 \leq j \leq n, (i,j) \neq (1,n)} \ell(\gamma^*_i) \geq \ell(\gamma^{n''}_s) + \sum_{1 \leq i \leq m, 1 \leq j \leq n, (i,j) \neq (1,n)} \ell(\gamma^{*'}_i),
\]
or equivalently
\[
\ell(\delta) - \ell(\gamma^{n1}_1) \geq \ell(\gamma^{n''}_s) + \ell(\delta^*) - \ell(\gamma^{n1'}_1).
\]
With inequality (3.4), we obtain
\[
\ell(\gamma^{n1'}_1) \geq \ell(\gamma^{n1}_1) + \ell(\gamma^{n''}_s). \tag{3.8}
\]
Using Corollary 2.2 we have
\[
\ell(\gamma^{n1}_1) + \ell(\gamma^{n''}_s) \geq 2\arcsinh \frac{\omega^{n1}_1 - 1}{2} + 2\arcsinh \frac{\omega^{n''}_s - 1}{30\hbar} \geq 2\arcsinh \frac{(\omega^{n1}_1 - 1)(\omega^{n''}_s - 1)}{30\hbar} \geq 2\arcsinh \frac{(\omega^{n1}_1 - 1)(\omega^*_s - 1)}{30\hbar}, \tag{3.9}
\]
and
\[
\ell(\gamma^{n1'}_1) \leq 2\arcsinh \frac{\omega^{n1'}_1}{2}.
\]
Recall that
\[
\omega^{n1'}_1 = (2n_1 - 1)\omega^1_1 + \cdots + \omega^{n1}_1 + \cdots + (2n_m - 1)\omega^1_m + \cdots + \omega^{n_m}_m + 1.
\]
and that $\omega_1^{n_1}$ and $\omega_*^*$ are the two biggest winding numbers of the strands in $S$. We then have

$$\omega_1^{n_1'} \leq \omega_1^{n_1} + (n_1^2 - 1)\omega_*^* + \cdots + n_m^2 \omega_*^* + 1$$

$$< \omega_1^{n_1} + (n_1^2 + n_2^2 + \cdots + n_m^2 - 1)\omega_*^* + \omega_*^*$$

$$\leq \omega_1^{n_1} + n^2 \omega_*^*.$$

Therefore,

$$\ell(\gamma_1^{n_1}) < 2 \arcsinh \frac{\omega_1^{n_1} + n^2 \omega_*^*}{2}.$$

(3.10)

From inequalities (3.9) and (3.10), we can rewrite inequality (3.8) in terms of winding numbers as follows:

$$2 \arcsinh \frac{\omega_1^{n_1} + n^2 \omega_*^*}{2} > 2 \arcsinh \frac{(\omega_1^{n_1} - 1)(\omega_*^* - 1)}{30h}.$$

Simplifying and rearranging yields

$$(16\hbar n^2 + 1)\omega_*^* - 1 > (\omega_*^* - 15\hbar - 1)\omega_1^{n_1}.$$

Since $\omega_*^* \geq 17\hbar$, we have

$$\omega_1^{n_1} \leq \frac{(16\hbar n^2 + 1)\omega_*^* - 1}{\omega_*^* - 15\hbar - 1} < \frac{16\hbar n^2 \omega_*^*}{\omega_*^* - 16\hbar} = \frac{16\hbar n^2}{\omega_*^* - 16\hbar} + \frac{16^2 \hbar^2 n^2}{\omega_*^* - 16\hbar}$$

$$\leq \frac{16\hbar n^2}{\omega_*^* - 16\hbar} + \frac{16^2 \hbar^2 n^2}{\hbar} = 272\hbar n^2.$$

(3.11)

**Bounding the self-intersection number of $S$ and finalizing the proof.** Notice that $\gamma_k$ contains $n$ geodesic segments entering cusp neighborhoods with boundary horocycles of length one. Each of them has to pass through an open cylinder bounded by the horocycle of length 2 and 1, and then return (see Figure 4). Such a cylinder is always embedded in $X$, and moreover, if $C$ and $C'$ are distinct cusps, then the associated cylinders are disjoint. Therefore,

$$\ell(\gamma_k) > n \times 2 \log(2) > n.$$

Since $\ell(\gamma_k) \leq C(k, X)$, it follows that

$$n < C(k, X).$$

(3.12)
Putting together inequalities (3.2), (3.3), (3.11) and (3.12), we have
\[ i(S, S) \leq (2n_1 - 1)\omega_1^1 + \cdots + \omega_1^{n_1} - n_1 + \cdots + (2n_m - 1)\omega_m^1 + \cdots + \omega_m^{n_m} - n_m \]
\[ < (n_1^2 + \cdots + n_m^2)\omega_1^{n_1} < n_1^2\omega_1^{n_1} < 272\kappa n^4 \leq 272\kappa (C(k, X))^4 \]
\[ < 272\kappa \left(\frac{\varepsilon}{\sqrt{h}}\right)^4 < \frac{k}{10}. \]
(3.13)

On the other hand, the \(\varepsilon_0\)-thick-thin decomposition of \(X\) decomposes \(\gamma_k\) into two parts, \(\gamma_{\varepsilon_0T} := \gamma_k \cap X_{\varepsilon_0T}\) and \(\gamma_{\varepsilon_0T} := \gamma_k \cap X_{\varepsilon_0T}\). By inequality (2.3), the self-intersection number of \(\gamma_{\varepsilon_0T}\) satisfies
\[ \ell(\gamma_{\varepsilon_0T}) > \frac{\varepsilon_0}{12} \sqrt{i(\gamma_{\varepsilon_0T}, \gamma_{\varepsilon_0T})}. \]
Using the fact that \(\ell(\gamma_{\varepsilon_0T}) < \ell(\gamma_k) \leq C(k, X)\), and inequality (3.2), we have
\[ i(\gamma_{\varepsilon_0T}, \gamma_{\varepsilon_0T}) < \left(\frac{12}{\varepsilon_0}\right)^2 (\ell(\gamma_k))^2 < \left(\frac{12}{\varepsilon_0}\right)^2 (C(k, X))^2 \]
\[ \leq \left(\frac{12}{\varepsilon_0}\right)^2 \left(\frac{\varepsilon}{\sqrt{h}}\right)^2 < \frac{k}{10}. \]
(3.14)

Note that by the choice of \(\varepsilon_0\), the self-intersection number that \(\gamma_k\) makes in \(U_{1 \leq i \leq m}B^i\) must be at least the number that \(\gamma_k\) makes in the closure of \(X_{\varepsilon_0T}\) (which means that we also count the self-intersection points of \(\gamma_k\) lying in \(\partial X_{\varepsilon_0T}\)). We therefore have
\[ i(S, S) \geq i(\gamma_k, \gamma_k) - i(\gamma_{\varepsilon_0T}, \gamma_{\varepsilon_0T}) > k - \frac{k}{10} = \frac{9k}{10}, \]
which contradicts to inequality (3.13). Thus, \(\omega_0^* < 17\kappa\), and we are done. \(\square\)

We now can prove the first part of Theorem 1.1

Proof of Theorem 1.1(1). By Lemma 3.1, for all \(k \geq K(X)\) and \(\gamma_k\) a shortest geodesic in \(\mathcal{G}_{\geq k}(X)\), \(\gamma_k\) enters the \(\varepsilon_0\)-thick-thin part of \(X\) only once. Let \(\mathcal{B} \subset X_{\varepsilon_0T}\) be the only cusp neighborhood with boundary a component of \(\partial X_{\varepsilon_0T}\) that \(\gamma_k\) enters. We can decompose \(\gamma_k\) into two parts, \(\gamma_{\varepsilon_0T} = X_{\varepsilon_0T} \cap \gamma_k\) and \(\gamma_{\varepsilon_0T} = X_{\varepsilon_0T} \cap \gamma_k\). We know that \(\gamma_{\varepsilon_0T}\) is a strand, so \(\gamma_{\varepsilon_0T}\) is also a continuous image of an interval. By inequality (3.14), the self-intersection number of \(\gamma_{\varepsilon_0T}\) is strictly smaller than \(k\). Now, suppose that we are at the initial point of \(\gamma_{\varepsilon_0T}\). By following \(\gamma_{\varepsilon_0T}\), we reach its terminal point, which is in \(\partial \mathcal{B}\). The only way we can continue is to go into the cusp neighborhood \(\mathcal{B}\) by following a strand. Since every loop we wrap around the cusp gives us one extra self-intersection, we will stop at the moment when we get enough \(k\) self-intersection points. Thus, \(i(\gamma_k, \gamma_k) = k\). \(\square\)

Remark 3.2. We can find the constant \(K(X)\) directly, meaning not via \(D(X)\), which implies that \(K(X)\) can be made much smaller, such that for all \(k \geq K(X)\), if \(\gamma_k \in \mathcal{G}_{\geq k}(X)\) is a minimal length geodesic, then \(\gamma_k\) self-intersects \(k\) times (see the proof of Theorem 1.1(1) above). However, by doing so, we do not know what \(\gamma_k\) looks like.

We now deal with the second part of Theorem 1.1. Since we want to be able to determine what all geodesics of minimal length in \(\mathcal{G}_{\geq k}(X)\) look like for \(k \geq K(X)\), we turn our attention to \(X_{\varepsilon_0T}\), the \(\varepsilon_0\)-thick part of \(X\).
Lemma 3.3. For all $k \geq K(X)$ and $\gamma_k$ a shortest geodesic in $G_{\geq k}(X)$, the self-intersection of $\gamma_{\varepsilon_0T} = X_{\varepsilon_0T} \cap \gamma_k$ satisfies

$$i(\gamma_{\varepsilon_0T}, \gamma_{\varepsilon_0T}) < D(X).$$

Recall that $D(X)$ is defined in (3.1). It is chosen in such a way that

$$\frac{\varepsilon_0}{12} \sqrt{x} > 2 \arcsinh \frac{\varepsilon_0 x}{2} + d(X) + 2 \varepsilon_0,$$

for all $x \geq D(X)$, where $d(X)$ is the maximum value of all shortest orthogonal distances from each boundary component of $X_{\varepsilon_0T}$ to itself.

Proof of Lemma 3.3. We construct a geodesic $\sigma \in G_{\geq k}(X)$, which by definition of $\gamma_k$, must not be shorter than $\gamma_k$. Our construction is based on the procedure used in [3, Proposition 4.2]. Let $B \subset X_{\varepsilon_0T}$ be as in the proof of Theorem 1.1(1), the only cusp neighborhood with boundary a component of $\partial X_{\varepsilon_0T}$ that $\gamma_k$ enters. Let $p, q \in \partial B$ be the initial and terminal points of $\gamma_{\varepsilon_0T}$. In our setting, $\delta_1$ has length at most $d(X)$, see the beginning of Section 3. After putting an orientation on $\delta_1$, denote its initial and terminal points by $x$ and $y$, respectively. Let $\delta_2$ be the shortest segment along $\partial B$ from $q$ to $x$, and $\delta_3$ be the shortest segment along $\partial B$ from $y$ to $p$. Each of $\delta_2$ and $\delta_3$ has length at most $\varepsilon_0$. Let $a$ be the point of $\gamma_{\varepsilon_0T}$ furthest away from $\partial B$, and $\alpha$ be the strand that has its endpoints at $a$ and winds around the cusp $i(\gamma_{\varepsilon_0T}, \gamma_{\varepsilon_0T})$ times. We orient $\alpha$ in the “winding” direction of $\gamma_{\varepsilon_0T}$. Now, define $\sigma$ to be the closed geodesic in the free homotopy class of the curve obtained by following $\gamma_{\varepsilon_0T}$ from $p$ to $a$, then following the strand $\alpha$, then continuing along $\gamma_{\varepsilon_0T}$ to $q$, and finally following $\delta_2 * \delta_1 * \delta_3$ to come back to $p$ (see Figure 5). By the construction, $\sigma \in G_{\geq k}(X)$, and hence $\ell(\gamma_k) \leq \ell(\sigma)$.

Figure 5. Constructing the geodesic $\sigma \in G_{\geq k}(X)$. 
We have
\[ \ell(\sigma) \leq \ell(\gamma_{\text{sel}}) + \ell(\alpha) + \ell(\delta_1) + \ell(\delta_2) \leq \ell(\gamma_{\text{sel}}) + \ell(\alpha) + d(X) + 2\epsilon_0. \]

Let \( h_a \) be the horocycle in the cusp neighborhood \( \mathcal{B} \) passing through \( a \). By abuse of notation, we denote the length of this horocycle also by \( h_a \). Note that by our setting, \( h_a \leq 2\epsilon < \epsilon_0 \).

Using Corollary 2.2, we have
\[ \ell(\alpha) \leq 2\arcsin \left( \frac{h_a i(\gamma_{\text{sel}}, \gamma_{\text{sel}})}{2} \right) < 2\arcsin \left( \frac{\epsilon_0 i(\gamma_{\text{sel}}, \gamma_{\text{sel}})}{2} \right). \]

Therefore,
\[ \ell(\sigma) < \ell(\gamma_{\text{sel}}) + 2\arcsin \left( \frac{\epsilon_0 i(\gamma_{\text{sel}}, \gamma_{\text{sel}})}{2} \right) + d(X) + 2\epsilon_0. \]

As \( \ell(\gamma_k) \leq \ell(\sigma) \), the above inequality also holds for \( \gamma_k \). By inequality (2.3), we have
\[ \ell(\gamma_k) = \ell(\gamma_{\text{sel}}) + \ell(\tau) > \ell(\gamma_{\text{sel}}) + \frac{\epsilon_0}{12} \sqrt{i(\gamma_{\text{sel}}, \gamma_{\text{sel}})}. \]

Thus,
\[ \frac{\epsilon_0}{12} \sqrt{i(\gamma_{\text{sel}}, \gamma_{\text{sel}})} < 2\arcsin \left( \frac{\epsilon_0 i(\gamma_{\text{sel}}, \gamma_{\text{sel}})}{2} \right) + d(X) + 2\epsilon_0. \]

By the choice of \( D(X) \), see [3.1], the above inequality does not hold if \( i(\gamma_{\text{sel}}, \gamma_{\text{sel}}) \geq D(X) \). Hence, \( i(\gamma_{\text{sel}}, \gamma_{\text{sel}}) < D(X) \) as desired. \( \square \)

We finally can complete the proof of Theorem 1.1.

**Proof of Theorem 1.1(2).** By Lemma 3.1, the geodesic \( \gamma_k \) can be decomposed into \( \gamma_{\text{sel}} \) and \( \gamma_{\text{sel}}T \), which are both continuous images of intervals. By Lemma 3.3, the self-intersection number of \( \gamma_{\text{sel}}T \) is strictly bounded from above by \( D(X) \). Let \( u \in X_{\text{sel}} \) be the self-intersection of \( \gamma_{\text{sel}} \) that is nearest to \( \partial \mathcal{B} \). Let \( u \) be the closed curve obtained by starting from \( u \), following \( \gamma_{\text{sel}} \) and returning to \( u \). Then \( u \) is freely homotopic to a power of \( \tau \), where \( \tau \) is a loop based at \( u \) and freely homotopic to the only cusp that \( \gamma_{\text{sel}}T \) winds around. Define \( \kappa := \gamma_k \setminus u \). We first note that the closed curve \( \kappa \) is not a power of \( \tau \) because \( \gamma_k \) is primitive. Besides, \( \kappa \) has the same self-intersection number as \( \gamma_{\text{sel}}T \), which is less than \( D(X) \).

If \( i(\gamma_{\text{sel}}T, \gamma_{\text{sel}}T) = i(\kappa, \kappa) = 0 \), then \( \gamma_k \) is freely homotopic to \( \kappa \ast u = \kappa \ast \tau^k \), where in this case, \( \kappa \) is a non-contractible simple loop based at \( u \) and not freely homotopic to the cusp. Thus, the theorem holds.

We now consider the case \( i(\gamma_{\text{sel}}T, \gamma_{\text{sel}}T) = i(\kappa, \kappa) \geq 1 \). By starting from its initial point, \( \gamma_{\text{sel}}T \) travels inside \( X_{\text{sel}} \) until it crosses itself for the first time. At this moment, it makes the first loop \( v \) and the first self-intersection point \( v \) (see Figure 6). More precisely, we choose a parametrization \( \gamma_{\text{sel}}T : [0,1] \to X \) and let \( t_2 \) be the supremum of all \( t \) such that the restriction \( \gamma_{\text{sel}}T|_{[0,t]} \) is a simple arc. Then there exists a unique \( t_1 \in [0,t_2] \) with \( \gamma_{\text{sel}}T(t_1) = \gamma_{\text{sel}}T(t_2) \), and it follows that \( \gamma_{\text{sel}}T|_{[t_1,t_2]} \) is a simple loop in \( \gamma_{\text{sel}}T \). This loop is denoted by \( v \) and the point \( \gamma_{\text{sel}}T(t_1) \) is denoted by \( v \). The simple loop \( v \) is non-contractible, so we consider whether it is freely homotopic to the cusp or not.

If \( v \) is freely homotopic to the cusp, we note that \( \gamma_{\text{sel}}T \) must go out of the cusp neighborhood with boundary \( v \), because otherwise, \( \kappa \) would only loop around the cusp, which contradicts the fact that \( \gamma_k \) is primitive. To get out of the cusp neighborhood, \( \gamma_{\text{sel}}T \) must create a bigon, which means that \( \gamma_k \) has excess self-intersections (see [10]), and this is absurd. Hence, \( v \) is not freely homotopic to the cusp.
Figure 6. If $v$ is freely homotopic to the cusp, then $\gamma_{\varepsilon_0 T}$ will create a bigon.

Define $\chi := \kappa \setminus v$. If $\chi$ is freely homotopic to a power of $\tau$, then $\gamma_k$ is freely homotopic to $v*\tau^k$, and hence the theorem holds.

Suppose that $\chi$ is not freely homotopic to a power of $\tau$. By observing that the loop $v \subset X_{\varepsilon_0 T}$ has length at least $2\varepsilon_0$ worth of length which we can use to wrap around the cusp, and since our $\varepsilon$ is chosen to be small enough, this adds as many self-intersections as we want.

Let $p, q, a, \alpha, h_a$ be as in the proof of Lemma 3.3, which are as follows: $p, q \in \partial B$ are the initial and terminal points of $\gamma_{\varepsilon_0 t}$, respectively, $a$ is the point of $\gamma_{\varepsilon_0 t}$ furthest away from $\partial B$, $\alpha$ is the strand that has its endpoints at $a$ and winds around the cusp $i(\gamma_{\varepsilon_0 T}, \gamma_{\varepsilon_0 T})$ times (in the same direction as the “winding” direction of $\gamma_{\varepsilon_0 t}$), $h_a$ is the horocycle passing through $a$ in the cusp neighborhood $B$, and by abuse of notation, we also denote by $h_a$ the length of this horocycle.

Let $\sigma$ be the closed geodesic in the free homotopy class of $\alpha * u * \chi$ (the curve obtained by following $\gamma_{\varepsilon_0 t}$ from $p$ to $a$, then following the strand $\alpha$, then continuing along $\gamma_{\varepsilon_0 T}$ until $v$, skipping the loop $\sigma$ to return immediately to $p$ by following $\gamma_{\varepsilon_0 T}$). By our construction, $\sigma \in \mathcal{G}_{\geq k}(X)$, and hence $\ell(\gamma_k) \leq \ell(\sigma)$. Using Corollary 2.2 and Lemma 3.3, we have

$$
\ell(\gamma_k) \leq \ell(\sigma) \leq \ell(\chi) + \ell(u) + \ell(\alpha) = \ell(\gamma_{\varepsilon_0 T}) + \ell(\gamma_{\varepsilon_0 t}) - \ell(v) + \ell(\alpha)
$$

$$
\leq \ell(\gamma_k) - 2\varepsilon_0 + 2 \text{arsinh} \left( \frac{h_a i(\gamma_{\varepsilon_0 T}, \gamma_{\varepsilon_0 T})}{2} \right)
$$

$$
\leq \ell(\gamma_k) - 2\varepsilon_0 + 2 \text{arsinh} \left( \frac{2\varepsilon D(X)}{2} \right)
$$

$$
\leq \ell(\gamma_k) - 2\varepsilon_0 + 2 \text{arsinh} \left( \frac{\varepsilon_0}{10D(X)} D(X) \right)
$$

$$
\leq \ell(\gamma_k) - 2\varepsilon_0 + 2 \text{arsinh} \left( \frac{\varepsilon_0}{10} \right)
$$

$$
< \ell(\gamma_k),
$$

which is a contradiction.

In conclusion, $\gamma_k$ is freely homotopic to $\kappa * \tau^k$, where $\tau$ is freely homotopic to the cusp, and $\kappa$ is a non-contractible simple loop that is not freely homotopic to $\tau$. \qed
4. Corollary and remarks

In this section, we will consider $\mathcal{M}_{g,n}$, the set of all cusped hyperbolic surfaces of genus $g$ with $n$ punctures or boundary components, that are not the thrice punctured sphere and the systole lengths of which are at least $s$, where $s > 0$. In order to prove Corollary 1.2, we will make all the constants defined at the beginning of Section 3 become constants that depend on $g, n$ and $s$. But before doing so, we first perform some computations related to pairs of pants with cusp(s).

**Lemma 4.1.** Let $Y$ be a pair of pants with at least one cusp. Let $d$ be the shortest orthogonal distance from the horocycle of length $h$ to itself.

1. If $Y$ has one cusp and two boundary geodesics of length $\alpha$ and $\beta$, then
   \[ d = d(h, \alpha, \beta) = 2 \log \frac{2(\cosh \frac{\alpha}{2} + \cosh \frac{\beta}{2})}{h}. \]

2. If $Y$ has two cusps and a boundary geodesic of length $\alpha$, then
   \[ d = d(h, \alpha) = 2 \log \frac{2(\cosh \frac{\alpha}{2} + 1)}{h}. \]

3. If $Y$ is the thrice punctured sphere, then
   \[ d = d(h) = 2 \log \frac{4}{h}. \]

![Figure 7](image-url)

**Proof.** Let $p_{\alpha\beta}$ be the geodesic segment that is perpendicular to both $\alpha$ and $\beta$. Upon lifting to the universal cover, we can normalize so that $p_{\alpha\beta}$ lifts to a geodesic segment that is contained on the Euclidean semicircle centered at the origin with radius one. The lifts of $\alpha$ and $\beta$ are geodesic segments $\tilde{\alpha}$ and $\tilde{\beta}$ that are contained on Euclidean semicircles orthogonal to the real axis. Denote the endpoints of $\tilde{\alpha}$ and $\tilde{\beta}$ by $x_\alpha, y_\alpha$ and $x_\beta, y_\beta$ respectively (see

---

1 The thrice punctured sphere will be treated in Remark 4.3.
Figure 7). Then the lift of the horocycle of length $h$ is the Euclidean line $y = \frac{x\beta + y\beta - x\alpha - y\alpha}{h}$.

Parametrizing $\tilde{\alpha}^2$ and $\tilde{\beta}^2$ as in the proof of Lemma 2.1, by computation

$$x_\alpha = -e^{-\frac{\alpha}{2}}, \quad y_\alpha = -e^{-\frac{\alpha}{2}}, \quad x_\beta = e^{-\frac{\beta}{2}}, \quad y_\beta = e^{\frac{\beta}{2}}.$$ 

Thus,

$$d(h,\alpha,\beta) = 2\log \frac{x_\beta + y_\beta - x_\alpha - y_\alpha}{h} = 2\log \frac{2(\cosh \frac{\alpha}{2} + \cosh \frac{\beta}{2})}{h},$$

proving the first part of the lemma. The remaining parts follow similarly. \(\Box\)

We now recall Bers’ theorem in the non-compact case (see [6, Theorem 5.2.6]). If $S$ is not the thrice punctured sphere, then there exists a pants decomposition of $S$ such that the simple closed geodesics are bounded above by

$$L(g,n) := 4(3g - 3 + n) \log \frac{4\pi(2g - 2 + n)}{3g - 3 + n}.$$

Using Bers’s constant and the previous lemma, we can find an upper bound for the shortest orthogonal distance from an embedded horocycle in $S$ to itself.

**Corollary 4.2.** The shortest orthogonal distance from an embedded horocycle of length $h$ in $S$ to itself is at most

$$d(h,g,n) = 2\log \frac{4\cosh \frac{L(g,n)}{2}}{h}.$$

**Proof.** This is a direct consequence of Lemma 4.1. \(\Box\)

We can now prove Corollary 1.2 by using the same reasoning as the one given in the proof of Theorem 1.1, with the constants depending on $g$, $n$ and $s$.

**Proof of Corollary 1.2.** Let $S \in \mathcal{M}_{g,n}^+$. A result by Adams [1] states that the length of the longest embedded horocycle in $S$ is bounded above by

$$\bar{h} = \bar{h}(g,n) := 12g + 5n - 11.$$

By our convention, the systole of $S$ is at least $s$. Set

$$\bar{h}_0 = \bar{h}_0(g,n) := \frac{1}{30\bar{h}},$$

and

$$\varepsilon_0 = \varepsilon_0(g,n,\bar{s}) := \min \left\{ \frac{\bar{h}_0}{5}, \frac{\bar{s}}{2} \right\}.$$

Denote by $S_{\varepsilon_0 T}$ the $\varepsilon_0$-thick part of $S$, and $S_{\varepsilon_0 t}$ the $\varepsilon_0$-thin part of $S$. By the choice of $\varepsilon_0$, the $\varepsilon_0$-thin part $S_{\varepsilon_0 t}$ is the union of embedded cusp neighborhoods. Each boundary component of $S_{\varepsilon_0 t}$ is indeed an embedded horocycle boundary of a cusp of length $\frac{2}{\sqrt{\cosh \varepsilon_0} - 1}$. We can find the shortest orthogonal distance from each boundary component of $S_{\varepsilon_0 t}$ to itself, which by Corollary 4.2 is at most

$$\mathcal{A}(g,n,\bar{s}) = 2\log \left( 2\sqrt{\cosh \varepsilon_0 - 1} \cosh \frac{L(g,n)}{2} \right).$$

Choose $D(h,g,\bar{s}) \geq 2$ such that

$$\frac{\varepsilon_0}{12} \sqrt{x} > 2\arcsinh \frac{\varepsilon_0 x}{2} + \mathcal{A}(g,n,\bar{s}) + 2\varepsilon_0, \quad \text{for all } x \geq D(h,g,\bar{s}).$$
Set
\[\varepsilon = \varepsilon(g, n, \delta) := \frac{\varepsilon_0}{10D(h, g, \delta)}.\]

Denote by \(S_{\varepsilon T}\) the \(\varepsilon\)-thick part of \(S\), and \(S_{\varepsilon t}\) the \(\varepsilon\)-thin part of \(S\). By Corollary 1.2, the shortest orthogonal distance from the length one horocycle boundary of a cusp to the horocycle itself is at most
\[d(g, n) = 2\log\left(4\cosh\frac{L(g, n)}{2}\right).\]

Choose \(K(g, n, \delta) \geq D(g, n, \delta)\) such that
\[C(k, g, n) := 2\arcsinh(k) + d(g, n) + 1 < \frac{\varepsilon}{h\sqrt{k}}, \quad \text{for all} \quad k \geq K(g, n, \delta).\]

Fix once and for all the constants \(\varepsilon, \varepsilon_0, h_0\), \(D(g, n, \delta)\), and \(K(g, n, \delta)\). For all \(k \geq K(g, n, \delta)\), we consider \(G_{\geq k}(S)\), the set of closed geodesics on \(S\) that self-intersect at least \(k\) times, and investigate those of minimal length. The procedure now is the same as in the proof of Theorem 1.1. □

In what follows, we show that the lower bound \(\delta\) on the systole lengths is necessary to make sure that the shortest geodesics are \(ba\) geodesics.

**Example 4.1.** Given \(g, n\), if there is no lower bound on systole lengths, we can construct a surface \(S\) of genus \(g\) with \(n\) cusps such that for \(k \geq 100\), shortest geodesics in \(G_{\geq k}(S)\) are not \(ba\) geodesics.

If \(g = 0\), our construction works for \(n \geq 9\). If \(g = 1\), it works for all \(n \geq 3\). If \(g \geq 2\), it works for all \(n \geq 1\). What we really use in the condition of \((g, n)\) is that, if we decompose \(S\) into pants, there is a pair of pants without cusp whose boundary lengths can be made as small as we want.

Before doing the construction, we first need some formulas. Let \(Y(a, b, c)\) be a pair of pants with boundary geodesics of lengths \(a, b\) and \(c\). Let \(\alpha\) and \(\beta\) be generators of the fundamental group of \(Y\) which wind counter-clockwise around two boundary geodesics of lengths \(a\) and \(b\) respectively. Consider the geodesic in the free homotopy class of \(\beta^{-1}\alpha^k\). Its length, which can be computed by using the crossed right-angled hyperbolic hexagon formula, is given by
\[\cosh\frac{\ell(\beta^{-1}\alpha^k)}{2} = \frac{\sinh k\frac{a}{2}}{\sinh \frac{c}{2}} \left(\cosh\frac{c}{2} + \cosh\frac{a}{2} \cosh\frac{b}{2}\right) + \cosh\frac{ka}{2} \cosh\frac{b}{2}.\]  

This geodesic is somewhat similar to \(ba\) geodesics for the case of cusp(s). In what follows, we will need to consider multiple geodesics of this type. Therefore, to simplify, we will call them \(back\) geodesics.

We will also need to use the Collar Theorem. It says that small geodesics have large tubular neighborhoods, which are topological cylinders [6, Theorem 4.1.1]. For a simple closed geodesic \(\gamma\) (of length \(\gamma\)) on \(S\), its collar is of width
\[\omega(\gamma) = \arcsinh\frac{1}{\sinh \frac{\gamma}{2}},\]
Now let $Y_1(x,x,x)$ be the pair of pants with boundary geodesics of the same length $x$. Let $g_1$ be a back geodesic of $Y_1$. By equation (4.1), its length is given by

$$\cosh \frac{\ell(g_1)}{2} = \sinh \frac{kx}{2} \left( \cosh \frac{x}{2} + \left( \cosh \frac{x}{2} \right)^2 \right) + \cosh \frac{kx}{2} \cosh \frac{x}{2}. \quad (4.3)$$

Let $Y_2(0,x,y)$ be the pair of pants with one cusp and two boundary geodesics of lengths $x,y$ ($x < y$). Let $g_2$ be the shortest $ba_c^k$ geodesic of $Y_2$, which is the one that winds around the cusp $k$ times and that winds around the boundary component of length $x$ once. Its length is given by

$$\cosh \frac{\ell(g_2)}{2} = k \left( \cosh \frac{x}{2} + \cosh \frac{y}{2} \right) + \cosh \frac{x}{2}.$$

Let $Y_3(0,0,y)$ be the pair of pants with two cusps and one boundary geodesic of length $y$. Let $g_3$ be the shortest $ba_c^k$ geodesic of $Y_3$, which is the one that winds $k$ times around one cusp and one time around the other cusp. Its length is given by

$$\cosh \frac{\ell(g_3)}{2} = k \left( \cosh \frac{y}{2} + 1 \right) + 1 < \cosh \frac{\ell(g_2)}{2}.$$

Set

$$y = 2 \arcsinh \frac{1}{\sinh \frac{k}{16}}, \quad x = \frac{y}{\sqrt{2k + 1}}.$$

With these values of $x$ and $y$, the following conditions hold for all $k \geq 100$,

- $\ell(g_1) < \ell(g_3)$,
- $\ell(g_1) < 2w(y)$.

We can now construct $S$. For $g = 0$, we first glue $Y_1$ to three copies of $Y_2$ along the boundary geodesics of length $x$, and then glue each $Y_2$ to a copy of $Y_3$ along the boundary geodesics of length $y$ (see Figure 8 for $S \in \mathcal{M}_{0,9}$). To obtain $S \in \mathcal{M}_{0,n}$ with $n \geq 10$, we can replace a copy of $Y_3$ by a surface $S'$ of genus 0 and $n - 7$ cusps with systole length $y$.

![Figure 8. $S \in \mathcal{M}_{0,9}$.](image)
For $g = 1$, we first glue two boundary components of $Y_1$ together to get a one-holed torus, then glue the one-holed torus to a copy of $Y_2$, along the boundary geodesic of length $x$. Finally, we glue the obtained surface to a surface $S'$, of genus 0 and $n-1$ cusps with systole length $y$, along the boundary geodesic of length $y$ (see Figure 9 for $S \in \mathcal{M}_{1,3}$).

![Figure 9. $S \in \mathcal{M}_{1,3}$.](image)

For $g \geq 2$, we first glue two boundary components of $Y_1$ together to get a one-holed torus, then glue the one-holed torus to $Y_2$, along the boundary geodesic of length $x$. Finally, we glue the obtained surface to a surface $S'$, of genus $g-1$ and $n-1$ cusps with systole length $y$, along the boundary geodesic of length $y$ (see Figure 10 for $S \in \mathcal{M}_{2,1}$).

![Figure 10. $S \in \mathcal{M}_{2,1}$.](image)

We now prove that for all $k \geq 100$, the geodesic $g_1$ is shorter than all $ba^k_c$ geodesics on $S$. Let $g$ be a $ba^k_c$ geodesic on $S$. We know that $g$ stays on a pair of pants with at least one cusp. If $g$ stays on a copy of $Y_2$, $Y_3$ or on $S'$, then
\[
\ell(g) \geq \ell(g_3) > \ell(g_1).
\]
Otherwise, $g$ must pass through a collar of width at least $w(y)$, and then return. Therefore,
\[
\ell(g) > 2w(y) > \ell(g_1).
\]
Thus, shortest geodesics in $\mathcal{G}_{\geq k}(S)$ are not $ba^k_c$ geodesics.

**Remark 4.3.** Denoting by $Y$ the thrice punctured sphere, for any non-negative integer $k$, we consider $\mathcal{G}_{\geq k}(Y)$, the set of all closed geodesics on $Y$ with at least $k$ self-intersection number. Let $\gamma_k$ be a shortest geodesic in $\mathcal{G}_{\geq k}(Y)$. For $k \in \{0, 1\}$, $\gamma_k$ self-intersects once. We want to know $i(\gamma_k, \gamma_k)$ for $k \notin \{0, 1\}$. By computing the constants defined in the proof of Theorem 1.1, we have $i(\gamma_k, \gamma_k) = k$ for all $k \geq 10^{35}$ (here we compute the constant $K(Y)$ directly, meaning not via $D(Y)$, see Remark 3.2 for more details).
For $2 \leq k < 10^{35}$, we might expect to find $i(\gamma_k, \gamma_k)$ by a combinatorial method. Recall that each closed geodesic on $Y$ determines a so-called word length, which is the number of letters required to express the geodesic as a reduced cyclic word in terms of the generators of the fundamental group and their inverses. The relation between the word length and the self-intersection number of a geodesic on the thrice punctured sphere has been investigated by Chas and Phillips [7]. Since we can put an upper bound on the hyperbolic length of the shortest geodesics in $\mathcal{F}_{\geq k}(Y)$, we can also put an upper bound on their word lengths. Thus, to find $i(\gamma_k, \gamma_k)$, the aim is to look at the set of all reduced cyclic words up to a certain word length, to subsequently use the Cohen - Lustig [8] algorithm in order to compute their self-intersection numbers, and to finally determine the shortest ones in $\mathcal{F}_{\geq k}(Y)$. However, since the number of involved words is tremendous, substantial computational power is needed, which makes it difficult to get examples.

ACKNOWLEDGEMENT

The author would like to thank Hugo Parlier for many valuable discussions and suggestions, as well as for his support and encouragement. Furthermore, she would also like to thank Binbin Xu for fruitful conversations and remarks.

REFERENCES

1. Colin Adams, *Maximal cusps, collars, and systoles in hyperbolic surfaces*, Indiana Univ. Math. J. 47 (1998), no. 2, 419–437. MR 1647904
2. Ara Basmajian, *The stable neighborhood theorem and lengths of closed geodesics*, Proc. Amer. Math. Soc. 119 (1993), no. 1, 217–224. MR 1152271
3. ______, *Universal length bounds for non-simple closed geodesics on hyperbolic surfaces*, J. Topol. 6 (2013), no. 2, 513–524. MR 3065183
4. ______, *Short geodesics on a hyperbolic surface*, Recent advances in mathematics, Ramanujan Math. Soc. Lect. Notes Ser., vol. 21, Ramanujan Math. Soc., Mysore, 2015, pp. 39–43. MR 3380365
5. Lipman Bers, *An inequality for Riemann surfaces*, Differential geometry and complex analysis, Springer, Berlin, 1985, pp. 87–93. MR 780038
6. Peter Buser, *Geometry and spectra of compact Riemann surfaces*, Modern Birkhäuser Classics, Birkhäuser Boston, Inc., Boston, MA, 2010, Reprint of the 1992 edition. MR 2742784
7. Moira Chas and Anthony Phillips, *Self-intersection numbers of curves in the doubly punctured plane*, Exp. Math. 21 (2012), no. 1, 26–37. MR 2904905
8. Marshall Cohen and Martin Lustig, *Paths of geodesics and geometric intersection numbers. I*, Combinatorial group theory and topology (Alta, Utah, 1984), Ann. of Math. Stud., vol. 111, Princeton Univ. Press, Princeton, NJ, 1987, pp. 479–500. MR 895629
9. V. Erlandsson and H. Parlier, *Short closed geodesics with self-intersections*, ArXiv e-prints (2016).
10. Joel Hass and Peter Scott, *Intersections of curves on surfaces*, Israel J. Math. 51 (1985), no. 1-2, 90–120. MR 804478

Department of Mathematics, University of Luxembourg, Luxembourg

E-mail address: thihan.vo@uni.lu