

The shortest non-simple closed geodesics on hyperbolic surfaces

Ara Basmajian*, Hugo Parlier[†] and Hanh Vo

Abstract. This article explores closed geodesics on hyperbolic surfaces. We show that, for sufficiently large k , the shortest closed geodesics with at least k self-intersections, taken among all hyperbolic surfaces, all lie on an ideal pair of pants and have length $2 \operatorname{arccosh}(2k + 1)$.

1. Introduction

The study of closed geodesics on hyperbolic surfaces has multiple facets which links together topics as diverse as spectral theory, symbolic dynamics, geometric topology and, of course, hyperbolic geometry. Without going into its rich history, in the past few decades, simple closed geodesics have played a capital role in the understanding of Teichmüller and moduli spaces, whereas closed geodesics were more traditionally related to spectral theory. The study of closed geodesics, and in particular the non-simple ones, has seen a recent resurgence over the last decade and a plethora of recent results suggest that stratifying closed geodesics by their self-intersection number, and asking similar questions to the ones that were asked for simple curves, provides fascinating directions of research.

One way to do this is to study the family of curves with fixed self-intersection. Notable examples of successes in that direction include generalizations of Mirzakhani's simple curve counting theorem [14] to curves with (fixed) self-intersection (see for instance [10, 11, 15, 18]).

In a somewhat orthogonal direction, there has been a lot of work relating how intersection and length are related, in particular when the intersection grows. For each surface and each natural number, there is a shortest closed geodesic with (at least) that number of self-intersections. These are a sort of generalization of the systole (the shortest closed geodesic on a surface), and their geometry and actual number of self-intersections was first systematically studied in [2]. While a lot of work has gone into extending these results [3, 4, 7, 9], to this point it has been unclear whether these shortest curves have exactly the minimum number of self-intersections. The most precise result to date [22] says that if a surface has a cusp, then for large enough k , the shortest geodesic with at least k self-intersections has exactly k self-intersections. As we will see, this result will provide a crucial step in our investigation.

Let γ_k be the closed geodesic on a connected hyperbolic surface X_k of minimal length among

*Supported by a grant from the Simons foundation (359956, A.B.) and PSC-CUNY Award 65245-00 53.

[†]Supported by the Luxembourg National Research Fund OPEN grant O19/13865598.

2020 *Mathematics Subject Classification*: Primary: 32G15. Secondary: 57K20, 30F60.

Key words and phrases: closed curves, hyperbolic surfaces, moduli spaces

all primitive closed geodesics with at least k self-intersection points taken among all closed geodesics on all possible hyperbolic surfaces. Our main result is the following.

Theorem 1.1. *For large enough k , X_k is an ideal pair of pants and γ_k is a corkscrew geodesic with exactly k self-intersections.*

By *corkscrew* geodesic we mean a geodesic in the homotopy class as described in Figure 1: that is a curve consisting of the concatenation of a simple arc and another that winds k -times around a boundary. We can also make the statement "for large enough k " explicit: our results hold for all $k \geq 10^{125}$. As one might imagine, we have not tried to optimize this lower bound.

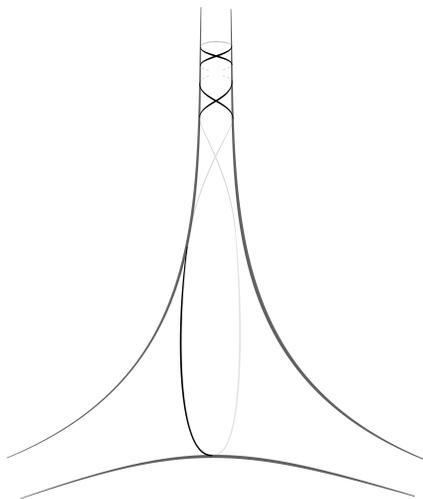


Figure 1: A corkscrew curve

The problem we solve in the above theorem was first introduced in [3] where such curves were shown to satisfy length bounds, namely upper and lower bounds that are roughly logarithmic in k . The upper bound follows from an explicit construction and the corkscrew geodesics above are exactly the ones provided in [3] and they have length exactly $2 \operatorname{arccosh}(2k + 1)$. They are the natural candidate curves that one might expect as they lie on the smallest possible hyperbolic surface. Our result shows that these candidate curves, at least for sufficiently large k , are indeed minimal among all possible geodesics with at least k self-intersections.

We also point out that, during the writing of this paper, Shen and Wang [19] posted a paper which shows that the asymptotic growth investigated in [3] is on the order of $2 \log(k)$. While our theorem implies theirs, the methods seem quite different.

One might ask what happens for smaller k . It is a result of Yamada [23] that the shortest non-simple geodesic taken among all surfaces is precisely on an ideal pair of pants and has 1 self-intersection (it is a so-called figure 8 geodesic which is a corkscrew geodesic with a single twist). More generally, Buser [5] showed that on any hyperbolic surface, the shortest

non-simple closed geodesic is always a figure 8 curve, from which one could also deduce the result for $k = 1$. Other than $k = 1$, no other explicit values were known before Theorem 1.1 above.

The proof is somewhat involved and requires multiple, but mostly elementary, ingredients. One of the main new ingredients is to decompose the surface into thick and thin parts when the notion of thick and thin depends on the intersection number. The whole point is to show that, for large enough k , the only optimal curve is the corkscrew geodesic on the ideal pair of pants. For this purpose, another essential ingredient is the use of the ideal pair of pants as a companion surface on which we construct a comparison curve.

The paper is structured as follows: after a preliminary section, we provide a sketch of proof which explains the different ingredients. The final section contains all the missing details.

Acknowledgements.

The authors thank a variety of people they have discussed these ideas with including Binbin Xu and Peter Buser.

2. Preliminaries

We consider orientable hyperbolic surfaces, that is surfaces endowed with a complete Riemannian metric of constant curvature -1 . For our purposes, it suffices to consider finite-type surfaces. In a slight abuse of notation, by boundary we mean the boundary of the Nielsen core of the surface (so the part of the surface where we've removed infinite funnels).

By curve we mean the continuous image of a circle which is not freely homotopic to a point or a cusp. We assume curves to be primitive, that is not powers of another curve, and not oriented. Generally we think of a curve as a free homotopy class of curve, and we make readily use of the fact that there is a unique geodesic in each free homotopy class.

The self-intersection of a curve is trickier to define than one might assume. A relatively easy way is to define it as the number of double points (counted with multiplicity) of a geodesic representative. Note that this only works because geodesics minimize intersection and we've assumed that our curves are primitive. We will sometimes refer to a closed geodesic with at least k self-intersection points as a k -geodesic. For instance, all closed geodesics are 0-geodesics.

Given a hyperbolic surface X and a point $p \in X$, the injectivity radius of X at p is the radius of the largest open disk embedded at p and will be denoted $\text{injr}_{ad}_X(p)$. Given a threshold radius $r > 0$, the surface X can be decomposed into its r -thin part (resp. thick part), corresponding to the set of points where the injectivity radius is at most (resp. at least) r .

With this in hand, we can now define the quantities we are interested in. Set

$$M_k := \inf\{\ell_Y(\alpha) \mid Y \text{ is a hyperbolic surface and } i(\alpha, \alpha) \geq k\}$$

We will generally denote by γ_k or γ the curve which we assume is optimal among all geodesics with at least k self-intersections taken among all geodesics on any hyperbolic surface. We denote by X_k or X the optimal surface on which γ lives. Observe that here we are assuming an existence result which follows from compactity results of completions of moduli spaces and which is explicitly proven in [3].

A useful tool, introduced by Thurston in [21], is as follows.

Theorem 2.1. *A surface Y with a boundary geodesic of positive length can be deformed to a surface with a cusp X and such that for all closed curves δ , the corresponding closed geodesics satisfy $\ell_X(\delta) < \ell_Y(\delta)$.*

We refer to [16, 17] and [8] for other proofs, extensions and applications.

As mentioned before, an essential ingredient in our approach is the following result of Vo [22].

Theorem 2.2. *For any $s_0 > 0$ and $\kappa_0 \in \mathbb{N}$ there exists a constant $K = K(s_0, \kappa_0)$ such that for any $k \geq K$, for any surface X with $\text{sys}(X) \geq s_0$ and $\kappa(X) \leq \kappa_0$, the shortest k -geodesic is a $b(a_c)^k$ geodesic.*

Vo uses the term $b(a_c)^k$ geodesics for curves that consist of an arc that winds around a cusp k times and an additional simple arc between the endpoints of the first arc. We use the term "ideal pair of pants" for the unique thrice-punctured sphere (a hyperbolic sphere with three cusps). Note that on an ideal pair of pants, the candidate curves are exactly $b(a_c)^k$ geodesics and have length $2 \operatorname{arccosh}(2k + 1)$. Furthermore, the above result can be quantified for the ideal pair of pants [22].

Corollary 2.3. *On the ideal pair of pants, for any $k > 10^{125}$, the shortest k -geodesic is a corkscrew geodesic with exactly k self-intersections.*

It is this result in particular that we need to prove our theorem.

3. Sketch of proof

Before proving the needed estimates, we provide a sketch of proof which gives the outline and strategy of our approach. One can imagine that a seasoned hyperbolic geometer would be able to fill in the details themselves, and can then skip the rest of the paper.

The approach requires dividing the surface into thick and thin parts, where what we mean by thick and thin will be carefully chosen. In what follows, k is always going to be our

self-intersection threshold which, at the level of our sketch proof, we can assume is as large as we need it to be at each step. There are finitely many steps.

Recall that our ultimate goal is to show that for large enough k , the only optimal geodesic and corresponding surface are on ideal pairs of pants. These geodesics have length exactly $2 \operatorname{arccosh}(2k + 1)$, which is asymptotic to $2 \log(k)$ for large k .

The curve γ fills X

A very useful technique is to use Theorem 2.1 introduced above. It shows that γ is a filling curve on X , and that $X \setminus \gamma$ is a collection of disks and once-punctured disks where, geometrically, the punctures are realized as cusps. This is because otherwise one can apply a stretch path to strictly reduce the length of all curves (including γ) which means that γ was not optimal.

Relating intersection and length in the thick part

A first key idea, which is already present in most papers relating length to intersection, is to observe that self-intersection of an arc c in the (uniformly) thick part of a surface (where thick is fixed) is quadratically related to its length. That is

$$\ell(c)^2 \gtrsim i(c, c)$$

where by \gtrsim we mean bigger than up to a positive multiplicative constant. So on such a surface, its length is roughly bounded below by the square root of its self-intersection number. Observe this immediately prevents, for large enough k , a surface which is uniformly thick to be optimal, simply because \sqrt{k} grows a lot faster than $\log(k)$.

This argument can be pushed further to investigate a type of relative thick part, where the thickness depends on k . Our measure of thickness will be of the type $\frac{1}{k^a}$ for a (fixed) positive a , generally between 0 and $\frac{1}{2}$. A key step is now to estimate how much length is necessary to go from a portion of a surface which is thick (of say injectivity radius 1) to the $\frac{1}{k^a}$ -thin part.

Being relatively thin

Being thin can take two forms: a thin part is either the collar of a short simple closed geodesic or it is the thin part of a horodisk neighborhood of a cusp. In either case, one can estimate that it takes at least $a \log(k) + O(1)$ length to reach the thin part. A precise statement, which works in both situations, is given in Lemma 4.1.

Now because we only have " $2 \log(k)$ " worth of length, for fixed a , this means that the curve cannot contain too many arcs that travel from thick to $\frac{1}{k^a}$ thin. In particular, the curve cannot travel to the $\frac{1}{k^a}$ thin part for $a > 1$, but we will be using another cut off point. A choice that will work for us will be to use $a = \frac{2}{5}$. It is somewhat arbitrary, but it has the important property and that the curve can contain at most 2 arcs that travel back and forth between the

thick and thin regions, and if there are indeed 2 of them, the arcs will be forced to travel even less into the very thin regions. By investigating the behavior of arcs in these relatively thin regions, this in turn will allow us to bound the intersection provided by these arcs, but before explaining how this works, we make a crucial observation.

Living in the $\frac{1}{k^{\frac{2}{5}}}$ -thick part

For k large enough, we consider the portion of γ that lives in the $\frac{1}{k^{\frac{2}{5}}}$ -thick part. We denote by X^T , where T is for Thick, the $\frac{1}{k^{\frac{2}{5}}}$ -thick part of the optimal surface X , and by X^t , where t stands for thin, the complement of X^T . Similarly, we denote γ^T the portion of γ that lies in X^T and by γ^t the portion of γ that lies in X^t .

We then show that γ^T cannot contribute significantly to intersection. Namely, for large enough k , we have

$$i(\gamma^T, \gamma^T) < k^{\frac{9}{10}}.$$

This is a consequence of Lemma 4.2. Again, there is a choice involved in the statement. Here we choose $\frac{9}{10}$ because it is bigger than 2 times $\frac{2}{5}$ but less than 1.

For simplicity, we use this technique throughout - that is to choose explicit constants slightly larger than what is strictly needed - to avoid having to deal with logarithmic terms and other lower order terms.

Now this means that γ^t must have (at least) self-intersection $k - k^{\frac{9}{10}}$. As we will show, this will force it to enter an even thinner region of the surface.

Thin collars about geodesics don't help

We begin by showing that on an optimal surface X , the $\frac{1}{k^{\frac{2}{5}}}$ -thin part X^t does not contain any simple closed geodesics.

To show this, first observe that if X^t did contain a simple closed geodesic then it would be of length less than $\frac{2}{k^{\frac{2}{5}}}$ and γ would have to cross it. In particular γ would contain a simple arc that goes from one end of the collar to the other which requires roughly $\frac{4}{5} \log(k)$ of length. An important subtlety here is that we are analyzing the length of the simple arc on X , not on X^t , although we only analyse intersection in X^t . Namely, the natural collar about the putative simple closed geodesic.

If there are no further arcs that enter the collar in X^t , then that portion of γ^t does not contribute any intersection and we continue to study the remainder of γ^t , but with less length at our disposal and we'll reach a contradiction later. So in order for there to be intersection inside the collar, as the initial arc is simple, there must be another arc that enters the collar.

This second arc is either a simple arc that traverses the collar or a non-simple arc that enters

and exits on the same side of the collar. We will refer to these two types of arcs as *transversal arcs* or *returning arcs* (see Figure 2).

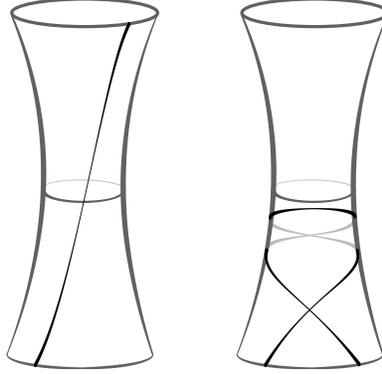


Figure 2: A transversal arc and a returning arc.

In both cases, this additional arc has to have length at least roughly $\frac{4}{5} \log(k)$. Note that this prevents a third arc from entering the collar in the thin region. Furthermore, we now have a better lower bound on the length of the core curve in the collar. The shorter the core is, the longer the transversal arc must be. And, by considering the length created by the second arc, the transversal curve can be at most on the order of $\frac{6}{5} \log(k)$. This gives a lower bound on the length of the core curve on the order of $\frac{1}{k^{\frac{3}{5}}}$, a considerable improvement over the previous lower bound which was on the order of $\frac{1}{k}$ (see Lemma 4.5 for the statement we use).

We can now look at the two arcs and, using the lower bound on the length of the core curve, give an upper bound on their intersection and self-intersection (Proposition 4.6). Essentially this boils down to the fact that in a collar, intersection and self-intersection is controlled linearly by the amount of times you wind around the core curve (a type of winding number). As the core curve has length at least on the order of $\frac{1}{k^{\frac{3}{5}}}$ and the total length of γ is at most on the order of $2 \log(k)$, this means that intersection is bounded above by a function on the order of $\log(k)k^{\frac{3}{5}}$. Again, this contributes (relatively) little to the needed k self-intersection even if we use up all the length.

All the action happens in the cusp

We can now safely assume that the majority of the intersection happens in the cusp, that is at least $k + o(k)$, and again we analyze the behavior in the cusp. Strands that enter a cusp region are necessarily returning strands and, by virtue of Lemma 4.1, if they reach the cusp region of X^t , then they have length at least on the order of $\frac{4}{5} \log(k)$. In particular, there can be at most two of them.

We first consider the case where there are two strands. By performing the same analysis as in the thin collar case, and for essentially the same reasons, we can conclude that the intersection

is bounded above by a function on the order of $\log(k)k^{\frac{3}{5}}$. We can thus conclude that there must be a single strand.

The amount of intersection provided by a single strand depends entirely on how high into the cusp the strand travels. As it must provide on the order of k self-intersection points, it must travel to the approximately $\frac{1}{k}$ thin part (see Lemma 5.3). This requires roughly $2 \log(k)$ length. Note that this estimate shows that (the length of the shortest k -geodesic) M_k has $2 \log(k)$ asymptotic growth (which is the main result of [19]).

The final blow

We are now in a position to put serious restrictions on what type of curves the γ_k can be. The fact that we have a strand that travels high into the cusp allows us to prove Proposition 5.4 which says that outside of the arc that travels high in the cusp, the rest of the curve has uniformly bounded length. In the statement we prove it for 5, but while we make use of the fact that it is quite short, the actual amount is not that important.

This proposition has an essential corollary: we can deduce that γ_k lies in an ideal pair of pants (Corollary 5.5). The corollary is proved using the isoperimetric inequality (in the hyperbolic plane and the once-punctured hyperbolic plane).

With this in hand, we are now in position to apply Vo's result (Corollary 2.3). It states that for all $k > 10^{125}$, the shortest k -geodesics on an ideal pair of pants are exactly our candidate curves and this completes the proof.

We now provide the details necessary for the above steps.

4. The structure of optimal geodesics for large enough k

One of the key aspects of our approach is to use the fact that it costs length to reach the thin part. This is quantified by the following lemma.

Lemma 4.1. *Fix a constant $N > e$ and a hyperbolic surface X .*

An arc c that goes from p to q with $\text{inrad}_X(p) = \text{arcsinh}(1)$ and $\text{inrad}_X(q) \leq \frac{1}{N}$ satisfies

$$\ell(c) > \log(N) - 1.$$

Proof. The point q lies in the half-collar of a curve of length at most $\frac{2}{N}$ (this length could be 0 in which case it lies in a horocycle neighborhood of a cusp).

If p doesn't lie in the same half-collar then any arc joining p and q must pass through another point p' with injectivity radius $\text{arcsinh}(1)$ so we may as well assume that p lies in the same half-collar as q . In this case the loops γ_p and γ_q of lengths $2 \text{arcsinh}(1)$ and less than $\frac{2}{N}$, based respectively at p and q , are isotopic. In this isotopy class there is a core closed geodesic α (or a

cusp which we can think of as a core geodesic of length 0). We may assume that the length of γ_q is exactly $\frac{2}{N}$ since any geodesic loop whose length is less than that is even further from p . We consider the cylinder C bounded by the loop γ_p based in p on one end and by α on the other. Consider δ the unique geodesic of shortest length between p and α (which is of infinite length if α is a cusp).

The point q belongs to the circle of points of this cylinder that are the base points of geodesics loops of length $\frac{2}{N}$. This circle of points is parallel to α (or is a horocycle in the case of a cusp). Of all points of this circle of points, the closest one to p is exactly the intersection between this circle and δ . Hence, as we want a lower bound on the distance between p and q , we can assume that q is exactly this point (see Figure 3).

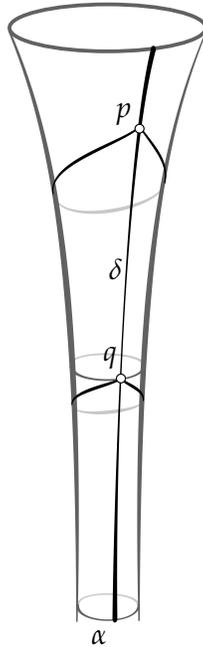


Figure 3: The distance between p and q is minimal when they both lie on δ

If we cut the cylinder C along δ , we obtain a quadrilateral or, in the case where we have a cusp, a triangle with an ideal vertex. It has an axis of reflection, say β , which cuts the quadrilateral into two quadrilaterals with two right angles as in Figure 4.

As q lies on δ , we can represent the loop γ_q on the quadrilateral and it too is invariant by the reflection. We denote by t the distance between p and q on the quadrilateral (and on the surface).

Our goal is to compute a lower bound for t using elementary hyperbolic geometry. Our first argument is geometric: as the lengths $\operatorname{arcsinh}(1)$ and $\frac{1}{N}$ are fixed, the geometry of

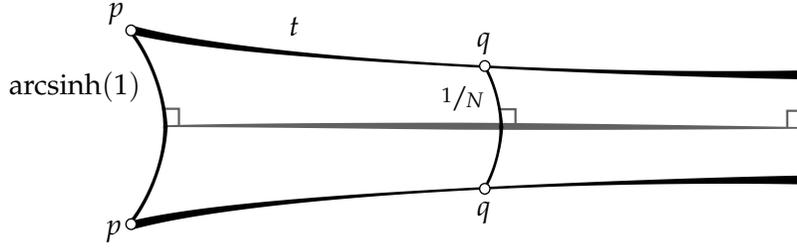


Figure 4: Quadrilaterals

the quadrilaterals, and hence t , is entirely determined by the length of the core geodesic α . Observe that t behaves monotonically in the length of α , and is shortest when $\ell(\alpha) = 0$. This can be seen geometrically - the arc of length $\frac{2}{N}$ has to move closer to p as the length of α decreases - or by a standard computation. In any event, as we are interested in a lower bound on t , we can assume the length of α is 0 and we are in the case where the quadrilateral has degenerated to an ideal triangle as in Figure 5.

It is fortuitous that in this case, the computations become easier. Using standard hyperbolic trigonometry, the (half)-angle at p is $\frac{\pi}{4}$ as in Figure 5.

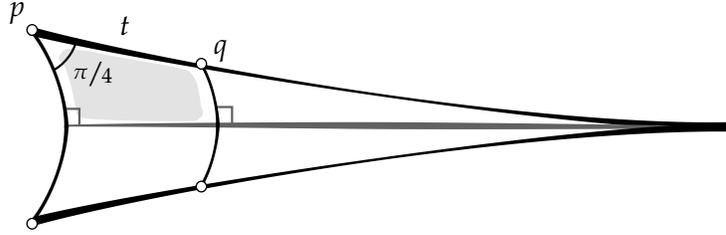


Figure 5: The worst case scenario is a triangle with an ideal vertex

Now we can compute t using the quadrilateral with two right angles and sides of length $\text{arcsinh}(1)$, $\frac{1}{N}$, t and angle at p equal to $\frac{\pi}{4}$ (shaded in Figure 5). Using the formula for such quadrilaterals (see for instance [13, Section VI.3, pp. 88-89]) we have

$$\sinh\left(\frac{1}{N}\right) = \sinh(\text{arcsinh}(1)) \cosh(t) - \cosh(\text{arcsinh}(1)) \sinh(t) \cos\left(\frac{\pi}{4}\right)$$

and thus

$$\sinh\left(\frac{1}{N}\right) = \cosh(t) - \sinh(t) = e^t.$$

From this we obtain

$$t = \log\left(\frac{1}{\sinh\left(\frac{1}{N}\right)}\right) > \log(N) - 1$$

for all $N > e$ (the last inequality can be verified by a quick calculus computation). This completes the proof. \square

Note that the constant 1 in the above statement is chosen for simplicity. The quantity

$$\log\left(\frac{1}{\sinh\left(\frac{1}{N}\right)}\right) - \log(N)$$

goes to 0 as N increases. The lower bound on N ensures that the loop around q is inside the half-collar bounded by the loop around p . Of course it is chosen so that the quantity $\log(N) - 1$ is positive.

For a real number $\frac{1}{2} > a > 0$ we provide a general upper bound on the amount of intersection that can occur in the $\frac{1}{k^a}$ -thick part of X . More precisely, consider $\gamma_a := \gamma \cap X^{\frac{1}{k^a}-T}$.

Lemma 4.2. $i(\gamma_a, \gamma_a) \leq (2k^a \log(k))^2$

Note that the statement is true for all $a > 0$ but as by definition the self-intersection of γ is k , for $a \geq \frac{1}{2}$ the lemma is void of content.

Proof. We break up γ_a into a maximal collection of segments of length at most $\frac{1}{k^a}$. Observe that any two of these segments can intersect at most once.

As $\ell(\gamma_a) \leq \ell(\gamma) \leq 2 \operatorname{arccosh}(1 + 2k) < 2 \log(k) + 4$, the maximal number of segments is

$$N \leq \left\lceil \frac{\ell(\gamma_a)}{\frac{1}{k^a}} \right\rceil < k^a(2 \log(k) + 4) + 1 < 2\sqrt{2}k^a \log(k)$$

where the last inequality holds for all $k > 10^6$.

And so the total number of intersections is $N(N-1)/2 < N^2/2$ and hence at most

$$\frac{1}{2}(2\sqrt{2}k^a \log(k))^2 = (2k^a \log(k))^2$$

as required. \square

Recall that the optimal k -geodesic γ is filling on the optimal surface X , that γ^T denotes the portion of γ in the $\frac{1}{k^{2/5}}$ thick part of X , and that we may assume there exists a point in X for which the injectivity radius is at most $\frac{1}{k^{2/5}}$.

Corollary 4.3. For large enough k , $i(\gamma^T, \gamma^T) < k^{\frac{9}{10}}$.

Proof. Setting $a = \frac{2}{5}$ in lemma 4.2, it suffices to observe that for large enough k ,

$$(2k^{\frac{2}{5}} \log(k))^2 < k^{\frac{9}{10}}.$$

It boils down to checking for which k we have

$$4(\log(k))^2 < k^{\frac{1}{10}}$$

which holds for all $k \geq 10^{48}$. □

We now turn to the part of γ in the $\frac{1}{k^{\frac{1}{5}}}$ -thin part denoted X^t . The following lemma shows that around collars of curves in X^t , the behavior of γ is well-controlled.

Lemma 4.4. *Suppose k is large. Let σ be a simple closed geodesic in X^t and C_σ be the collar region of σ in X^t . Then $\gamma \cap C_\sigma$ consists of at most two arcs.*

Proof. This is more or less a direct consequence of Lemma 4.1, where $N = k^{\frac{2}{5}}$.

Indeed all strands inside C_σ must reach and leave the $\frac{1}{k^{\frac{1}{5}}}$ -thin region. There is a somewhat subtle point here: we're looking at the length of the extension of the arc that enters the thin part of the collar. These full arcs, by Lemma 4.1 above, are of length at least

$$2 \left(\frac{2}{5} \log(k) - 1 \right).$$

Now if there were at least 3 then the total length of γ would be at least

$$3 \left(\frac{4}{5} \log(k) - 2 \right) = \frac{12}{5} \log(k) - 6 > 2 \log(k) + 4 > \ell(\gamma)$$

which provides a contradiction if $\log(k) > 25$, that is if $k > e^{25}$. □

Using the lemma above, we can now show that if there is intersection in the collar of a curve in X^t , then the core curve cannot be too short.

Lemma 4.5. *Suppose k is large. Let σ be a simple closed geodesic in X^t . Denote by C_σ the collar region of σ in X^t , and by γ_σ the restriction of γ to C_σ . If $i(\gamma_\sigma, \gamma_\sigma) > 0$, then*

$$\ell(\sigma) > \frac{1}{k^{4/5}}.$$

Proof. As γ is filling, there must be at least one transversal strand in C_σ . Hence any intersection must come from the intersection between the transversal strand and the returning strand, or from the intersection between two transversal strands. As we suppose that $i(\gamma_\sigma, \gamma_\sigma) > 0$, there are at least 2 strands (and by the previous lemma, there are exactly 2).

As each strand has length at least $\frac{4}{5} \log(k) - 2$, this means that a transversal strand has length at most

$$2 \log(k) + 4 - \left(\frac{4}{5} \log(k) - 2 \right) = \frac{6}{5} \log(k) + 6.$$

Now if the core geodesic is of length less than $\frac{1}{k^a}$ then by Lemma 4.1, the length of the transversal strand is at least $2 \log(k^a) - 2$. Hence for any $a > \frac{3}{5}$, we would reach a contradiction for large enough k . We choose $a = \frac{4}{5}$ for simplicity and argue by contradiction. We have

$$\frac{6}{5} \log(k) + 6 > \frac{8}{5} \log(k) - 2$$

that is

$$8 > \frac{2}{5} \log(k)$$

which fails for $k > e^{20}$ and hence $\ell(\sigma) > \frac{1}{k^{\frac{4}{5}}}$. \square

We now bound the total intersection that occurs in collars of simple closed geodesics inside X^t . We denote by $\sigma_1, \dots, \sigma_n$ the collection of simple closed geodesics of X^t and by C the collection of their collars (so $C = \cup_{i=1}^n C_{\sigma_i}$). We denote by γ_C the restriction of γ to C (so $\gamma_C = \cup_{i=1}^n C_{\gamma_i}$).

Proposition 4.6. *For large enough k , there are at most 2 simple closed geodesics in X^t . Moreover:*

- if there are none, then X must contain a cusp.
- if there is one, then $i(\gamma_C, \gamma_C) < k^{\frac{9}{10}}$,
- if there are two, then $i(\gamma_C, \gamma_C) = 0$.

Proof. Note that there are at most two strands in total that lie in C , and as γ must fill there are at most two simple closed geodesics in X^t .

Suppose X^t has no simple closed geodesics. If X does not contain a cusp neighborhood, then $X = X^T$, $\gamma = \gamma^T$, and hence by Corollary 4.3, $i(\gamma, \gamma) < k^{\frac{9}{10}}$, leading to a contradiction since $i(\gamma, \gamma) \geq k$.

If X^t contains two simple closed geodesics then there must be two transversal strands to each of these geodesics, and hence $i(\gamma_C, \gamma_C) = 0$.

We are left to prove the case where there is one simple closed geodesic in X^t . Again if there is one transversal then $i(\gamma_C, \gamma_C) = 0$. Thus the only non-trivial case to consider is when there is one simple closed geodesic in X^t and two strands, one of which must be transversal. There are cases to consider:

Case 1: Both strands are transversal.

Denote the two strands a and b and assume that $i(a, b) > 1$. Note that both are individually simple and go from one end of the collar to the other "monotonically": that is, if we parametrized the collar with respect to the core geodesic and distance to the core geodesic, this distance would be monotonic. The main observation is that two consecutive intersection

points are joined by one arc of a and one arc of b . These are not homotopic arcs, and their concatenation is freely homotopic to the core geodesic. Hence it follows that

$$(i(a, b) - 1)\ell(\sigma) < \ell(a) + \ell(b)$$

(Note that we need at least 2 intersection points for this to work and each extra intersection point adds a new copy of a curve homotopic to σ .)

As $\ell(a) + \ell(b) < \ell(\gamma) < 2\log(k) + 4$ and by lemma 4.5, $\ell(\sigma) > \frac{1}{k^{4/5}}$, we have

$$i(a, b) < 1 + k^{4/5}(2\log(k) + 4) < k^{9/10}.$$

The last inequality holds for large enough k .

Case 2: There is one transversal and one returning strand.

Let a be the transversal strand and b the returning strand. The proof works almost identically as in case (1) but with two small changes.

The first one is that we have to account for the self-intersection of b . The second one is that b does not pass through the collar monotonically, but first approaches the core curve before turning back. To overcome the second difficulty, we split b into two simple and monotonic arcs b^+ and b^- .

Now we have $i(b, b) = i(b^+, b^-)$. As before, between consecutive intersection points, we have arcs of b^+ and b^- that can be concatenated to be homotopic to σ . Hence

$$(i(b, b) - 1)\ell(\sigma) < \ell(b^+) + \ell(b^-) = \ell(b).$$

With the same argument, we can bound the intersection between the arc a and b^+ , resp. b^- :

$$(i(a, b^+) - 1)\ell(\sigma) < \ell(a) + \ell(b^+).$$

and

$$(i(a, b^-) - 1)\ell(\sigma) < \ell(a) + \ell(b^-).$$

so

$$i(a, b) \leq 2 + \frac{\ell(a) + \ell(b)}{\ell(\sigma)}.$$

As before, we use the upper bound on the length of γ and the lower bound on the length of σ to obtain:

$$i(b, b) < 1 + k^{4/5}(2\log(k) + 4)$$

and

$$i(a, b) < 2 + k^{4/5}(2\log(k) + 4)$$

Hence

$$i(\gamma_C, \gamma_C) = i(a, b) + i(b, b) < 3 + 2k^{\frac{4}{5}}(2\log(k) + 4) < k^{\frac{9}{10}}$$

where the last inequality hold for all $k > 10^4$. \square

With these results in hand, for large enough k , observe that outside of cusp regions, γ self-intersects at most $2k^{\frac{9}{10}}$ times. Hence there must be $k - 2k^{\frac{9}{10}}$ intersections in cusp regions. In particular,

Corollary 4.7. *For large enough k , the optimal surface X must have a cusp.*

5. Optimal excursions into cusped surfaces and the main result

We now prove a general estimate that relates length to intersection in a cusp region. Similar estimates can be found in the literature [4] but here we follow [22].

Lemma 5.1. *Let a be a returning strand that enters and exits a cusp region at points on the horocycle boundary of length 1 and whose highest point has injectivity radius $\frac{1}{N}$ for some $N > 2$. Then*

$$2\sqrt{\frac{1}{4\sinh^2\frac{1}{N}} - 1} - 2 \leq i(a, a) \leq 2\sqrt{\frac{1}{4\sinh^2\frac{1}{N}} - 1} - 1 \leq N - 1.$$

Similarly, if a is as above and b is another one which is not shorter than a , then

$$i(a, b) \leq 4\sqrt{\frac{1}{4\sinh^2\frac{1}{N}} - 1} \leq 2N.$$

Proof. Upon lifting to the upper-half plane model for the hyperbolic plane, we can normalize so that the horocycle boundary of length 1 lifts to the Euclidean line $y = 1$ and the parabolic element identified with the element in the fundamental group that wraps once around the cusp is $f : z \mapsto z + 1$. The lift of a is a geodesic segment \tilde{a} that is contained on a Euclidean semicircle orthogonal to the real axis. This semicircle can be further normalized so that its center is the origin. Denote the endpoints of this semicircle by $-r$ and r . A computation yields

$$r = \cosh \frac{\ell(a)}{2}.$$

The f -translates of \tilde{a} intersects \tilde{a} exactly $i(a, a)$ times. We have

$$i(a, a) + 1 = \max \left\{ n \in \mathbb{N} : -r + n \leq -r + 2\sqrt{r^2 - 1} \right\} = \left\lfloor 2\sqrt{r^2 - 1} \right\rfloor.$$

Note that a point on the cusp region that has injectivity radius x belong to the horocycle of length $2\sinh(x)$. Since the highest point of a has injectivity radius $\frac{1}{N}$, we have

$$r = \frac{1}{2\sinh\frac{1}{N}}.$$

Thus

$$2\sqrt{\frac{1}{4\sinh^2\frac{1}{N}} - 1} - 2 \leq i(a, a) \leq 2\sqrt{\frac{1}{4\sinh^2\frac{1}{N}} - 1} - 1 \leq N - 1,$$

where the last inequality follows from $\sinh(x) > x$ for $x > 0$.

For two returning strands a and b , where a is as above, and b is not shorter, we have

$$i(a, b) \leq 2\lceil w(a) \rceil \leq 2(i(a, a) + 1),$$

where $w(\cdot)$ denotes the winding number and the first inequality follows from [9, Lemma 3.2].

And so using the above inequality we have

$$i(a, b) \leq 4\sqrt{\frac{1}{4\sinh^2\frac{1}{N}} - 1} \leq 2N.$$

□

We now show that γ will have at most 1 strand in cusp regions.

Lemma 5.2. *There is at most one strand of γ that enters a cusp neighborhood of X^t .*

Proof. The arguments are similar to the earlier arguments regarding strands of γ entering collar neighborhoods of the thin part. As before, using Lemma 4.1 there are at most 2 strands in total in X^t , so in particular in a cusp region of X^t . Now if there are exactly 2, then neither can go too high. More precisely, each strand has length at least $\frac{4}{5}\log(k) - 2$, and this means that the other strand has length at most

$$2\log(k) + 4 - \left(\frac{4}{5}\log(k) - 2\right) < \frac{6}{5}\log(k) + 6.$$

Hence, again by Lemma 4.1, if it enters the $\frac{1}{k^a}$ -thin region, then it is of length at least

$$2a\log(k) - 2.$$

So if $a \geq \frac{4}{5}$ then

$$\frac{8}{5}\log(k) - 2 < \frac{6}{5}\log(k) + 6$$

and thus

$$\frac{1}{5}\log(k) < 4$$

which is false for all $k > e^{20}$. Thus, for large enough k , the arcs never reach a point of injectivity radius $\frac{1}{k^{4/5}}$. We can now apply Lemma 5.1, their intersection number is bounded from above by $2k^{4/5}$. Hence, if there are 2 strands, for large enough k they cannot provide the needed intersection. □

Now that we know that we have a single strand in the cusp region, and that it must contribute at least $k - 2k^{\frac{9}{10}}$ to intersection, we have the following.

Lemma 5.3. *For large enough k , the curve γ contains a strand that reaches the $\frac{1}{k-2k^{\frac{9}{10}}}$ thin part of a cusp region.*

Proof. As before, we denote the strand by a and the minimal injectivity radius of the strand $\frac{1}{N}$. We can again apply Lemma 5.1, we have

$$i(a, a) < N$$

We know that $i(a, a) \geq k - 2k^{\frac{9}{10}}$ which proves the lemma. \square

At this point we have shown that the optimal k -geodesic γ (for k large) enters a single cusp in the X^t part of the optimal k -surface X and must reach the $\frac{1}{k-2k^{\frac{9}{10}}}$ thin part of this cusp. In addition, there are more than $k - 2k^{\frac{9}{10}}$ intersections in this cusp part.

We now show that outside of the arc a which we've described above, the rest of the curve is of length at most 5. The choice of 5 is somewhat arbitrary and we're not trying to optimize it, but it is small enough to suit our purposes.

Proposition 5.4. *For large enough k , the curve γ consists of a single arc a which travels into the cusp and a second arc b of length $\ell(b) < 5$*

Proof. We argue by contradiction. Suppose that the complementary arc of a is of length at least 5. We shall construct a geodesic γ' on the thrice punctured sphere of length shorter than γ and with at least as many intersection points.

Consider the arc a^t , the restriction of a to X^t . For large enough k , its complementary arc (that is $\gamma \setminus a^t$) has at most $2k^{\frac{9}{10}}$ self-intersections: at most $k^{\frac{9}{10}}$ in X^T (Corollary 4.3) and at most $k^{\frac{9}{10}}$ in the collars around curves in the thin part (Proposition 4.6). So a^t (and thus a) must have self-intersection at least $k - 2k^{\frac{9}{10}}$.

We now construct the geodesic γ' on the thrice punctured sphere as follows. We begin by taking a copy of a on a cusp of our choosing. Note that there are many ways to take a copy of a (by rotating any given copy) but we don't mind which one we choose here. We add $2k^{\frac{9}{10}} + 1$ intersection points to a to obtain an arc a' by adding extra winding by following the shortest geodesic loop based at the highest point. By Lemma 5.3, this shortest geodesic loop has length

$$L \leq \frac{2}{k - 2k^{\frac{9}{10}}} < \frac{1}{2k^{\frac{9}{10}} + 1}$$

for $k > 10^{10}$, so the length of a' is at most

$$L(2k^{\frac{9}{10}} + 1) + \ell(a) < \ell(a) + 1.$$

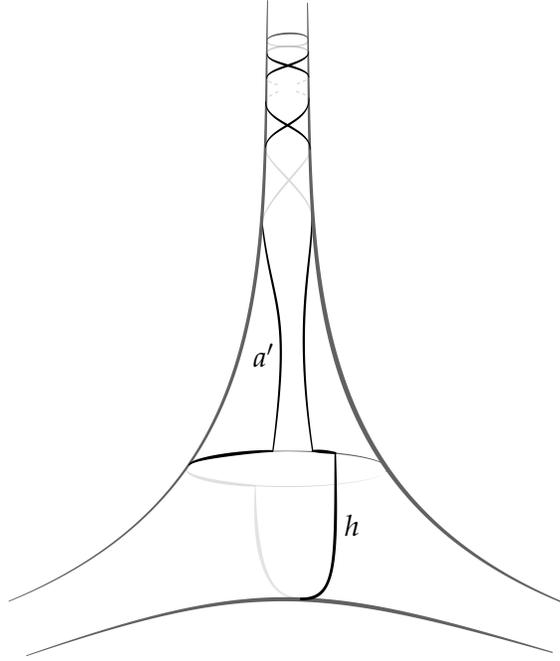


Figure 6: Constructing γ'

We now join the endpoints of a' in the following way. We use the shortest orthogeodesic h on the ideal pair of pants with endpoints on the horocycle of length 1, and add paths that follow the horocycle boundary and remain disjoint (see Figure 6). This extra arc is of length at most 1 plus the length of h which by a routine calculation on the ideal pants is at most $4 \log(2)$. So the total length of γ' is strictly less than $\ell(a) + 1 + 1 + 4 \log(2) < \ell(a) + 5$. Moreover, γ' has (at least) one extra intersection point compared to γ which we have supposed is of length $\geq 5 + \ell(a)$.

Now, though the curve we are considering is not geodesic, its homotopy class is clear, and so we control the self-intersection of the curve very well here. Finally, its geodesic representative will have length less than that of γ , leading to a contradiction. \square

Corollary 5.5. *For large enough k , the surface X is a pair of pants.*

Proof. We use the isoperimetric inequalities of the hyperbolic plane and the once-punctured hyperbolic plane, see, e.g., [1]. The former says that a rectifiable simple curve of length L that surrounds a region of area A in the hyperbolic plane satisfies

$$A^2 + 4\pi A \leq L^2.$$

The latter says that a rectifiable simple curve of length L that surrounds a once-punctured

(that is once-cusped) hyperbolic region of area A satisfies

$$A \leq L.$$

Note that the latter inequality is implied by the former.

Now we cut along the arc b which is of length < 5 by Proposition 5.4, and along the horocycle of length 1 around the cusp. The result is a collection of disks or once-punctured disks. As each curve contributes boundary length to exactly two regions, the total length ℓ of the boundaries of the disks is strictly less than $2(5 + 1) = 12$. Now for each disk, punctured or not, if ℓ_i is the boundary length and its area is a_i , then it satisfies $a_i \leq \ell_i$. By adding them up we have that the area of the surface satisfies

$$\text{Area}(X) \leq \ell < 12.$$

Now

$$\text{Area}(X) = -2\pi\chi(X) < 12$$

where $\chi(X)$ is the Euler characteristic of X , and hence

$$-\chi(X) = 2g - 2 + n < \frac{6}{\pi} < 2$$

and as it is an integer we have that $-\chi(X) = 1$ and X is thus a pair of pants. As it has only cusps as ends, it is an ideal pair of pants as required. \square

We can now deduce the main theorem. We recall its statement.

Theorem 5.6. *Let γ_k be the closed geodesic on a connected hyperbolic surface X_k of minimal length among all primitive closed geodesics with at least k self-intersection points taken among all closed geodesics on all possible hyperbolic surfaces. Then for large enough k , X_k is an ideal pair of pants and γ_k is a corkscrew geodesic with exactly k self-intersections.*

Proof. By the above (Corollary 5.5), $X = X_k$ is an ideal pair of pants. Now by Vo's theorem [22], for all $k \geq 10^{125}$, the shortest geodesic on an ideal pair of pants with at least k self-intersections is necessarily a corkscrew geodesic with exactly k self-intersections and we are done. \square

References

- [1] Adams, Colin and Morgan, Frank. Isoperimetric curves on hyperbolic surfaces. Proc. Amer. Math. Soc. 127 (1999), no. 5, 1347–1356.
- [2] Basmajian, Ara. The stable neighborhood theorem and lengths of closed geodesics. Proc. Amer. Math. Soc. 119 (1993), no. 1, 217–224.

- [3] Basmajian, Ara. Universal length bounds for non-simple closed geodesics on hyperbolic surfaces. *J. Topol.* 6 (2013), no. 2, 513–524.
- [4] Basmajian, Ara. Short geodesics on a hyperbolic surface. *Recent advances in mathematics*, 39–43, Ramanujan Math. Soc. Lect. Notes Ser., 21, Ramanujan Math. Soc., Mysore, 2015.
- [5] Buser, Peter. *Geometry and spectra of compact Riemann surfaces*. Reprint of the 1992 edition. Modern Birkhäuser Classics. Birkhäuser Boston, Ltd., Boston, MA, 2010.
- [6] Chas, Moira. Self-intersection numbers of length-equivalent curves on surfaces. *Exp. Math.* 23 (2014), no. 3, 271–276.
- [7] Chas, Moira. Relations between word length, hyperbolic length and self-intersection number of curves on surfaces. *Recent advances in mathematics*, 45–75, Ramanujan Math. Soc. Lect. Notes Ser., 21, Ramanujan Math. Soc., Mysore, 2015.
- [8] Danciger, Jeffrey, Guéritaud, François and Kassel, Fanny. Margulis spacetimes via the arc complex. *Invent. Math.* 204 (2016), no. 1, 133–193.
- [9] Erlandsson, Viveka and Parlier, Hugo. Short closed geodesics with self-intersections. *Math. Proc. Cambridge Philos. Soc.* 169 (2020), no. 3, 623–638.
- [10] Erlandsson, Viveka, Parlier, Hugo and Souto, Juan. Counting curves, and the stable length of currents. *J. Eur. Math. Soc. (JEMS)* 22 (2020), no. 6, 1675–1702.
- [11] Erlandsson, Viveka and Souto, Juan. Counting curves in hyperbolic surfaces. *Geom. Funct. Anal.* 26 (2016), no. 3, 729–777.
- [12] Erlandsson, Viveka and Souto, Juan. *Mirzakhani’s curve counting*. Forthcoming book.
- [13] Fenchel, Werner. *Elementary geometry in hyperbolic space*. With an editorial by Heinz Bauer. De Gruyter Studies in Mathematics, 11. Walter de Gruyter & Co., Berlin, 1989.
- [14] Mirzakhani, Maryam. Growth of the number of simple closed geodesics on hyperbolic surfaces. *Ann. of Math. (2)* 168 (2008), no. 1, 97–125.
- [15] Mirzakhani, Maryam. Counting Mapping Class group orbits on hyperbolic surfaces, arxiv preprint, 2016.
- [16] Papadopoulos, Athanase and Th eret, Guillaume. Shortening all the simple closed geodesics on surfaces with boundary. *Proc. Amer. Math. Soc.* 138 (2010), no. 5, 1775–1784.
- [17] Parlier, Hugo. Lengths of geodesics on Riemann surfaces with boundary. *Ann. Acad. Sci. Fenn. Math.* 30 (2005), no. 2, 227–236.

- [18] Sapir, Jenya. Bounds on the number of non-simple closed geodesics on a surface. *Int. Math. Res. Not. IMRN* 2016, no. 24, 7499–7545.
- [19] Shen, Wujie and Wang, Jiajun. Minimal length of nonsimple closed geodesics on hyperbolic surfaces, *Arxiv preprint*, 2022.
- [20] Tan, Ser Peow. Self-intersections of curves on surfaces. *Geom. Dedicata* 62 (1996), no. 2, 209–225.
- [21] Thurston, Willam. Stretch maps on surfaces, *Arxiv preprint*, 1998.
- [22] Vo, Hanh. Short closed geodesics on cusped hyperbolic surfaces. *Pacific J. Math.* (2022), 127–151.
- [23] Yamada, Akira. On Marden’s universal constant of Fuchsian groups. II. *J. Analyse Math.* 41 (1982), 234–248.

Addresses:

The Graduate Center, City University of New York, NY, New York, 10016 and Hunter College,
City University of New York, 695 Park Ave., New York, NY, 10065
DMATH, FSTM, University of Luxembourg, Esch-sur-Alzette, Luxembourg
School of Mathematical and Statistical Sciences, Arizona State University, USA

Emails:

ABasmajian@gc.cuny.edu
hugo.parlier@uni.lu
thihanhv@asu.edu