

Self-intersection numbers of curves in the doubly-punctured plane

Moira Chas and Anthony Phillips

September 17, 2019

Abstract We address the problem of computing bounds for the self-intersection number (the minimum number of self-intersection points) of members of a free homotopy class of curves in the doubly-punctured plane as a function of their combinatorial length L ; this is the number of letters required for a minimal description of the class in terms of a set of standard generators of the fundamental group, and their inverses. We prove that the self-intersection number is bounded above by $L^2/4 + L/2 - 1$, and that when L is even, this bound is sharp; in that case there are exactly four distinct classes attaining that bound. For odd L we conjecture a smaller upper bound, $(L^2 - 1)/4$, and establish it in certain cases, where we show it is sharp. Furthermore, for the doubly-punctured plane, these self-intersection numbers are bounded *below*, by $L/2 - 1$ if L is even, $(L - 1)/2$ if L is odd; these bounds are sharp.

1 Introduction

By the doubly-punctured plane we refer to the compact surface with boundary (familarly known as the “pair of pants”) obtained by removing, from a closed two-dimensional disc, two disjoint open discs. This work extends to that surface the research reported in [7] for the punctured torus. The doubly-punctured plane has fundamental group free on two generators; given a basis, say a, b , a free homotopy class of curves on the surface can be uniquely represented as a reduced cyclic word in the symbols a, b, A, B (where A stands for a^{-1} and B for b^{-1}). A *cyclic word* w is an equivalence class of words related by a cyclic permutation of their letters; we will write $w = \langle r_1 r_2 \dots r_n \rangle$ where the r_i are the letters of the word, and $\langle r_1 r_2 \dots r_n \rangle = \langle r_2 \dots r_n r_1 \rangle$, etc. *Reduced* means that the cyclic word contains no juxtapositions of a with A , or b with B . The *length* (with respect to the basis (a, b)) of a free homotopy class of

Key words and phrases: doubly-punctured plane, thrice-punctured sphere, pair of pants, free homotopy classes of curves, self-intersection.

2000 *Mathematics Subject Classification:* Primary 57M05, Secondary 57N50, 30F99.

curves is the number of letters occurring in the corresponding reduced cyclic word. This length is independent of the basis chosen, as long as that basis is *standard*, in the sense that its two elements can be represented by non-self-intersecting curves.

This work studies the relation between length (with respect to a standard basis) and the *self-intersection number* of a free homotopy class of curves: the smallest number of self-intersections among all general-position curves in the class. (General position in this context means as usual that there are no tangencies or multiple intersections). The self-intersection number is a property of the free homotopy class and hence of the corresponding reduced cyclic word w ; we denote it by $SI(w)$. Note that a word and its inverse have the same self-intersection number.

- Theorem 1.1.** (1) *The self-intersection number for a reduced cyclic word of length L on the doubly-punctured plane is bounded above by $L^2/4 + L/2 - 1$.*
- (2) *If L is even, this bound is sharp: for $L \geq 4$ and even, the cyclic words realizing the maximal self-intersection number are (see Figure 1) $(aB)^{L/2}$ and $(Ab)^{L/2}$. For $L = 2$, they are aa, AA, bb, BB, aB and Ab .*
- (3) *If L is odd, the maximal self-intersection number of words of length L is at least $(L^2 - 1)/4$.*

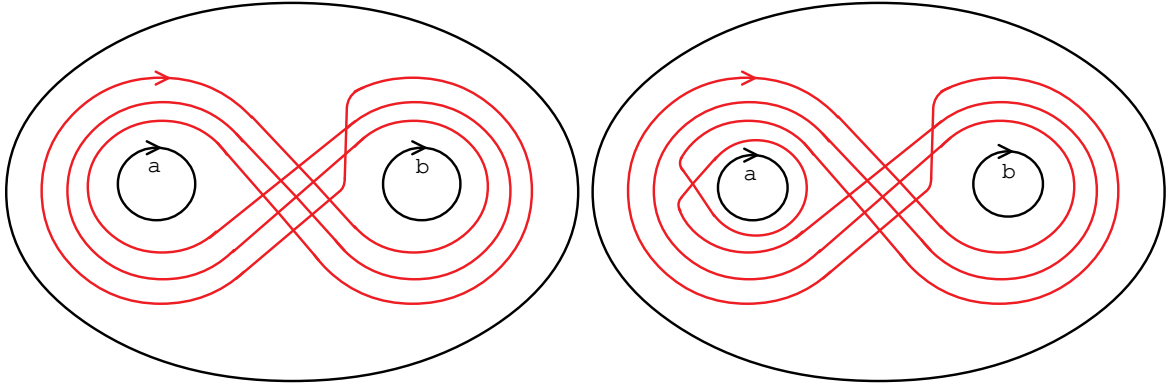


Figure 1: Left: curves of the form $\langle aBaBaB \rangle$ have maximum self-intersection number $L^2/4 + L/2 - 1$ for their length (Theorem 1.1). Right: curves of the form $\langle aaBaBaB \rangle$ have self-intersection number $(L^2 - 1)/4$; we conjecture (Conjecture 1.2) this is maximal, and prove this conjecture in certain cases (Theorem 1.4).

Conjecture 1.2. The maximal self-intersection number for a reduced cyclic word of odd length $L = 2k + 1$ on the doubly-punctured plane is $(L^2 - 1)/4$; the words realizing the maximum have one of the four forms $\langle (aB)^k B \rangle, \langle a(aB)^k \rangle, \langle (Ab)^k b \rangle, \langle A(Ab)^k \rangle$. ■

Definition 1.3. Any reduced cyclic word is either a pure power or may be written in the form $\langle \alpha_1^{a_1} \beta_1^{b_1} \dots \alpha_n^{a_n} \beta_n^{b_n} \rangle$, where $\alpha_i \in \{a, A\}$, $\beta_i \in \{b, B\}$, all a_i and b_i are positive, and $\sum_1^n (a_i + b_i) = L$, the length of the word. We will refer to each $\alpha_i^{a_i} \beta_i^{b_i}$ as an $\alpha\beta$ -block, and to n as the word's *number of $\alpha\beta$ -blocks*. ■

Theorem 1.4. *On the doubly-punctured plane, consider a reduced cyclic word w of odd length L with n $\alpha\beta$ -blocks. If $L > 3n$, or n is prime, or n is a power of 2, then the self-intersection number of w satisfies $\text{SI}(w) \leq \frac{L^2-1}{4}$. This bound is sharp.*

It is elementary to show that the only simple closed curves on the doubly punctured plane correspond to the empty word and the words a, b, ab and their inverses. (The doubly punctured plane is singular among surfaces of negative Euler characteristic in this respect, since every other such surface admits infinitely many distinct homotopy classes of simple curves [12]). This generalizes to the statement that, on the doubly punctured plane, self-intersection numbers of words are bounded *below*.

Theorem 1.5. *On the doubly punctured plane, curves in the free homotopy class represented by a reduced cyclic word of length L have at least $L/2 - 1$ self-intersections if L is even and $(L - 1)/2$ self-intersections if L is odd. These bounds are achieved by $(ab)^{\frac{L}{2}}$ and $(AB)^{\frac{L}{2}}$ if L is even and by the four words $a(ab)^{\frac{L-1}{2}}$, etc. when L is odd.*

Corollary 1.6. *A curve with minimal self-intersection number k has combinatorial length at most $2k + 2$. There are therefore only finitely many free homotopy classes with minimal self-intersection number k .*

Remark 1.7. A surface of negative Euler characteristic which is not the doubly punctured plane has infinitely many homotopy classes of simple closed curves [12]. Since the $(k + 1)$ st power of a simple closed curve has self-intersection number k , it follows that for any k there are infinitely many distinct homotopy classes of curves with self-intersection number k . (A more elaborate argument using the mapping class group constructs, for any k , infinitely many distinct *primitive* classes (not a proper power of another class) with self-intersection number k). So the doubly punctured plane is the unique surface of negative Euler characteristic satisfying Corollary 1.6. ■

Like our work in [7], the results presented in this article were suggested by experimental data generated by a JAVA program [6] based on the Cohen-Lustig algorithm [8]. The authors have benefited from discussions with Dennis Sullivan, and are very grateful to Igor Rivin who contributed an essential element to the proof of Theorem 1.5. Additionally, they have profited from use of Chris Arettines' JAVA program, which draws minimally self-intersecting representatives of free homotopy classes of curves in surfaces. The program is currently available [1].

1.1 Questions and related results

A free homotopy class of combinatorial length L in a surface with boundary can be represented by L chords in a fundamental polygon. Hence the maximal self-intersection number of a cyclic reduced word of length L is bounded above by $\frac{L(L-1)}{2}$. One may ask how closely the maximum can approach that bound.

We prove in [7] that for the punctured torus the maximal self-intersection number $\text{SI}_{\max}(L)$ of a free homotopy class of combinatorial length L is equal to $(L^2 - 1)/4$ if L

is even and to $(L-1)(L-3)/4$ if L is odd. This implies that the limit of $\text{SI}_{\max}(L)/L^2$ is $\frac{1}{4}$ as L approaches infinity. (Compare [10]). The same limit holds for the doubly punctured plane (Theorem 1.1). Our (limited) experiments do not suggest analogous polynomials for more general surfaces; but they do lead us to the following conjecture:

Conjecture 1.8. Consider closed curves on a surface-with-boundary S of Euler characteristic χ . Let $\text{SI}_{\max}(L)$ be the maximum self-intersection number for all curves on S of combinatorial length L . Then $\lim_{L \rightarrow \infty} \frac{\text{SI}_{\max}(L)}{L^2} = \frac{\chi}{2\chi-1}$. In particular, this limit approaches $\frac{1}{2}$ as the Euler characteristic of the surface approaches infinity. ■

The doubly punctured plane admits a hyperbolic metric making its boundary geodesic. An elementary argument shows that for curves on that surface, hyperbolic and combinatorial lengths are quasi-isometric. Some of our combinatorial results can be related in this way to statements about intersection numbers and hyperbolic length.

Basmajian [2] proved for a closed, hyperbolic surface S that there exists an increasing sequence M_k (for $k = 1, 2, 3, \dots$) going to infinity so that if w is a closed geodesic with self-intersection number k , then its geometric length is larger than M_k . Thus the length of a closed geodesic gets arbitrarily large as its self-intersection gets large. For the doubly punctured plane, in terms of the combinatorial length, we calculate $M_k = \sqrt{5 + 4k} - 1$.

Question 1.9. Consider closed curves on a hyperbolic surface S (possibly closed). Let $\text{SI}_{\max}(\ell)$ be the maximum self-intersection number for any curve of *hyperbolic* length at most ℓ . Does $\text{SI}_{\max}(\ell)/\ell^2$ converge? And if so, to what limit?

2 A linear model

In this section we will need to distinguish between a cyclically reduced linear word w in the generators and their inverses, and the the associated reduced cyclic word w . We introduce an algorithm for constructing from w a representative curve for w . An upper bound for the self-intersection numbers of these representatives may be easily estimated; taking the minimum of this bound over cyclic permutations of $\alpha\beta$ -blocks will yield a useful upper bound for $\text{SI}(w)$.

2.1 Skeleton words

Given a cyclically reduced word $w = \langle \alpha_1^{a_1} \beta_1^{b_1} \dots \alpha_n^{a_n} \beta_n^{b_n} \rangle$, where $\alpha_i = a$ or A , $\beta_i = b$ or B and all $a_i, b_i > 0$, the corresponding *skeleton word* is $w_S = \langle \alpha_1 \beta_1 \dots \alpha_n \beta_n \rangle$, a word of length $2n$.

We now describe a systematic way for drawing a representative curve for w_S starting from one of its linear forms w_S , and for *thickening* this curve to a representative for w .

The skeleton-construction algorithm: (See Figures 2 and 3) Start by marking off n points along each of the edges of the fundamental domain; corresponding points on the a, A sides are numbered $1, 3, 5, \dots, 2n - 1$ starting from their common corner; and similarly corresponding points on the b, B sides are numbered $2n, \dots, 6, 4, 2$, the numbers decreasing away from the common corner.

If the first letter in w_S is a , draw a curve segment entering the a -side at 1, and one exiting the A -side at 1 (vice-versa if the first letter is A). That segment is then extended to enter the b -side at 2 and exit the B -side at 2 if the next letter in w_S is b ; vice-versa if it is B . And so forth until the curve segment exiting the b (or B)-side at $2n$ joins up with the initial curve segment drawn.

We will refer to a segment of type ab, ba, AB, BA as a *corner segment*, and one of type aB, Ab, bA, Ba as a *transversal*. Note that (as above) a skeleton word has even length $2n$ and therefore has $2n$ segments (counting the *bridging segment* made up of the last letter and the first). The number of transversals must also be even, since if they are counted consecutively they go from lower-case to upper-case or vice-versa, and the sequence (upper, lower, ...) must end up where it starts. It follows that the number of corners is also even.

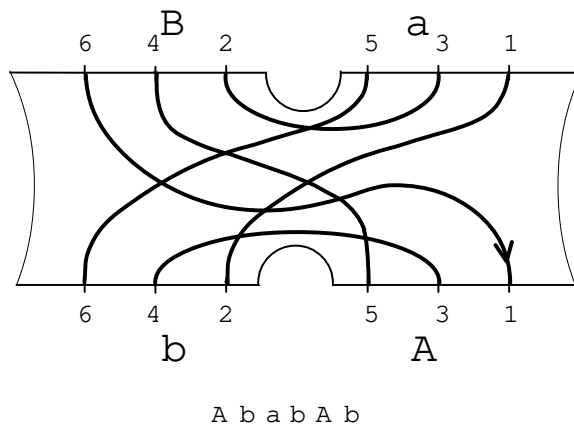


Figure 2: The skeleton curve $AbabAb$.

Proposition 2.1. *The self-intersection number of the representative of $(Ab)^n$ or $(aB)^n$ given by the curve-construction algorithm equals $n^2 + n - 1$.*

Proof. Consider $(Ab)^n$; see Figure 3, left. This curve has only transversals. There are n parallel segments of type Ab ; they join $1, 3, \dots, (2n - 1)$ on the a -side to $2, 4, \dots, 2n$ on the b -side. There are $n - 1$ parallel segments of type bA , which join $2, 4, \dots, 2n - 2$ on the B -side to $3, 5, \dots, 2n - 1$ on the A -side. Each of these intersects all n of the Ab segments. Finally the bridging bA segment joins $2n$ on the B -side to 1 on the A -side. This segment begins to the left of all the other segments and ends up on their right: it intersects all

$2n - 1$ of them. The total number of intersections is $n(n - 1) + 2n - 1 = n^2 + n - 1$. A symmetrical argument handles $(aB)^n$. ■

Proposition 2.2. *The self-intersection number of the representative of $(ab)^n$ given by the curve-construction algorithm equals $(n - 1)^2$.*

Proof. (See Figure 3, right) This curve has only corners. There are n segments of type ab , joining $1, 3, \dots, 2n - 1$ on the A -side to $2, 4, \dots, 2n$ on the b -side. Since their endpoints interleave, each of these curves intersects all the others. There are $n - 1$ segments of type ba , joining $2, 4, \dots, 2n - 2$ on the B -side to $3, 5, \dots, 2n - 1$ on the a -side. Again, each of these curves intersects all the others. Finally the bridging ba segment joining $2n$ to 1 spans both endpoints of all the others and so intersects none of them. The total number of intersections is $\frac{1}{2}n(n - 1) + \frac{1}{2}(n - 1)(n - 2) = (n - 1)^2$. ■

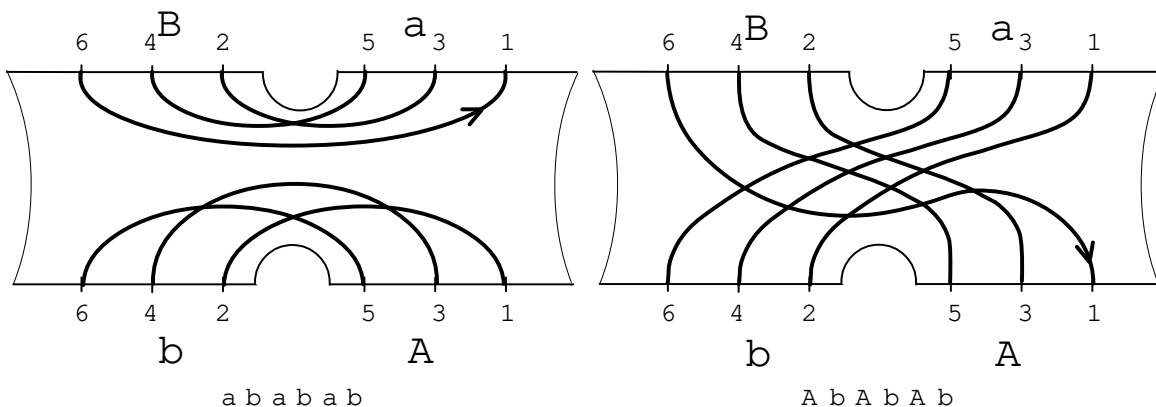


Figure 3: The skeleton curves $ababab$ and $AbAbAb$.

Proposition 2.3. *Let w be a skeleton word of length $2n$. The number of corner segments in w is even, as remarked above; write it as $2c$. Then the self-intersection number of w is bounded above by $n^2 + n - 1 - 2c$.*

Proof. Using Propositions 2.1 and 2.2 we can assume that w has both corner-segments and transversals. We may then choose a linear representative w with the property that the bridging segment between the end of the word and the beginning is a transversal. Of the $2c$ corners, c will be on top (those of type AB or ba) and c on the bottom (types ab and BA). An ab or AB corner segment joins a point numbered $2j - 1$ to a point numbered $2j$ on the same side, top or bottom, as $2j - 1$. It encloses segment endpoints $2j + 1, 2j + 3, \dots, 2n - 1, 2, 4, \dots, 2j - 2$, a total of $n - 1$ endpoints; similarly, a ba or BA segment encloses $n - 2$ endpoints. So there are at most $2c(n - 1) - c(c - 1)$ intersections involving corners, correcting for same-side corners having been counted twice. The $2n - 2c$ transversals intersect each other just as in the pure-transversal

case, producing $(n - c)^2 + (n - c) - 1$ intersections. The total number of intersections is therefore bounded by $n^2 + n - 1 - 2c$. Figure 2 shows the curve $AbabAb$ (here $n = 3, c = 1$) with 8 self-intersections. ■

2.2 Thickening a skeleton; proof of Theorem 1.1 (1), (2)

Once the skeleton curve corresponding to w_S is constructed, it may be *thickened* to produce a representative curve for w . The algorithm runs as follows.

The skeleton-thickening algorithm. (See Figure 4) Suppose for explicitness that w starts with A^{a_1} . The extra $a_1 - 1$ copies of A , inserted after the first one, correspond to segments entering the a -side (the first one at 1) and exiting the A -side (the last one at a point opposite the displaced entrance point of the first skeleton segment); the new segments are parallel. Similarly the extra $b_1 - 1$ segments appear as parallel segments originating and ending near the 2 marks on the b and B -sides; so there are no intersections between these segments and those in the first band. Proceeding in this manner we introduce n non-intersecting bands of $a_1 - 1, b_1 - 1, a_2 - 1, \dots, b_n - 1$ parallel segments. New intersections occur between these bands and segments of the skeleton curve. The two outmost bands (corresponding to a_1 and b_n) are each intersected by one of the skeleton segments; the next inner bands (a_2 and b_{n-1}) each intersect three of the skeleton segments; \dots ; the two innermost bands (a_n and b_1) each intersect $(2n - 1)$ of the skeleton segments.

Adding these intersections to the bound on the self-intersections of the skeleton curve itself yields

$$\text{SI}(w) \leq (a_1 + b_n - 2) + 3(a_2 + b_{n-1} - 2) + \dots + (2n - 1)(a_n + b_1 - 2) + n^2 + n - 1.$$

Since $1 + 3 + \dots + (2n - 1) = n^2$ we may repackage this expression as

$$\text{SI}(w) \leq f(a_1, \dots, a_n, b_1, \dots, b_n) - n^2 + n - 1,$$

where we define f by

$$f(a_1, \dots, a_n, b_1, \dots, b_n) = (a_1 + b_n) + 3(a_2 + b_{n-1}) + \dots + (2n - 1)(a_n + b_1).$$

Applying the skeleton-thickening algorithm to the cyclic permutation $\alpha_1^{a_1} \beta_1^{b_1} \dots \alpha_n^{a_n} \beta_n^{b_n} \rightarrow \alpha_2^{a_2} \beta_2^{b_2} \dots \alpha_n^{a_n} \beta_n^{b_n} \alpha_1^{a_1} \beta_1^{b_1}$ yields another curve representing the same word. There are n such permutations, leading to

$$(1) \quad \text{SI}(w) \leq \left[\min_{i=0, \dots, n-1} f \circ r^i(a_1, \dots, a_n, b_1, \dots, b_n) \right] - n^2 + n - 1,$$

where r is the coordinate permutation $(a_1, \dots, a_n, b_1, \dots, b_n) \rightarrow (a_2, \dots, a_n, a_1, b_2, \dots, b_n, b_1)$.

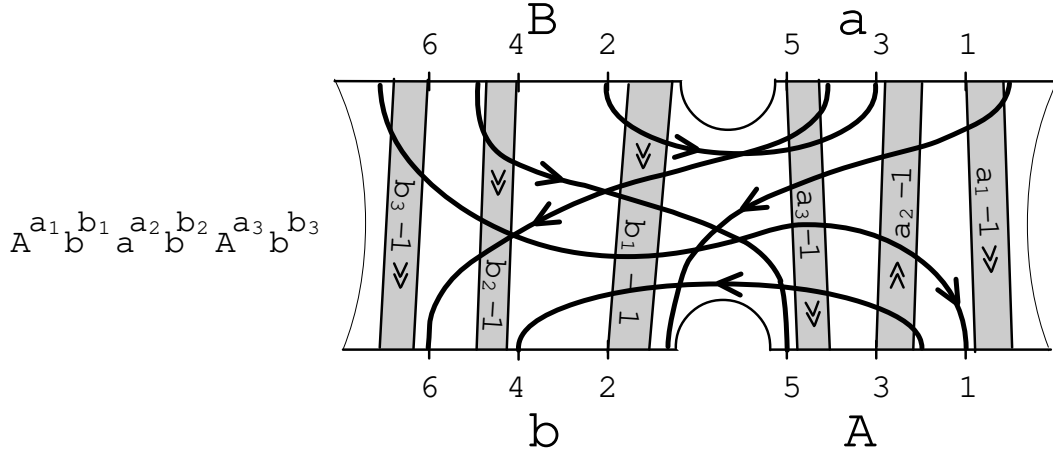


Figure 4: The skeleton curve $AbabAb$ thickened to represent the linear word $A^{a_1}b^{b_1}A^{a_2}b^{b_2}A^{a_3}b^{b_3}$. The grey bands represent the curve segments corresponding to the extra letters: $a_1 - 1$ copies of A , etc. Notice that the segments from the skeleton curve intersect the a_1 and b_3 bands once, the a_2 and b_2 bands three times, and the a_3 and b_1 bands five times.

Proposition 2.4. *Set $L = a_1 + \dots + a_n + b_1 + \dots + b_n$. Then $\min_{i=0, \dots, n-1} f \circ r^i(a_1, \dots, a_n, b_1, \dots, b_n) \leq nL$.*

Proof. We write

$$f(a_1, \dots, b_n) = (a_1 + b_n) + 3(a_2 + b_{n-1}) + \dots + (2n - 1)(a_n + b_1)$$

$$f \circ r(a_1, \dots, b_n) = (a_2 + b_1) + 3(a_3 + b_n) + \dots + (2n - 1)(a_1 + b_2)$$

⋮
⋮
⋮

$$f \circ r^{n-1}(a_1, \dots, b_n) = (a_n + b_{n-1}) + 3(a_1 + b_{n-2}) + \dots + (2n - 1)(a_{n-1} + b_n).$$

The average of these n functions is $\frac{1}{n}(L + 3L + \dots + (2n - 1)L) = nL$. Since the minimum of n functions must be less than their average, the proposition follows. ■

Proof of Theorem 1.1, (1) and (2) We work with $w = \langle \alpha_1^{a_1} \beta_1^{b_1} \dots \alpha_n^{a_n} \beta_n^{b_n} \rangle$. We have established that

$$\text{SI}(w) \leq \min_{i=0, \dots, n-1} f \circ r^i(a_1, \dots, a_n, b_1, \dots, b_n) - n^2 + n - 1.$$

Using Proposition 2.4,

$$\text{SI}(w) \leq nL - n^2 + n - 1 = -n^2 + n(L + 1) - 1.$$

For a given L , this function has its real maximum at $n = (L + 1)/2$. Since each $\alpha\beta$ -block contains at least 2 letters, n must be less than or equal to $L/2$. So a bound on $\text{SI}(w)$ is the value at $n = L/2$ (L even) or $n = (L - 1)/2$ (L odd):

$$\text{SI}(w) \leq \begin{cases} L^2/4 + L/2 - 1 & (L \text{ even}) \\ L^2/4 + L/2 - 7/4 & (L \text{ odd}). \end{cases}$$

For L even, note (Proposition 2.1) that the skeleton words $w = (aB)^n$ and $w = (Ab)^n$ satisfy $\text{SI}(w) = n^2 + n - 1 = L^2/4 + L/2 - 1$; so the bound for this case is sharp; furthermore since words with $n = L/2$ must be skeleton words, it follows from Proposition 2.3 these are the only words attaining the bound. ■

Remark 2.5. For L odd, our numerical experiments (which go up to $L = 20$) and the special cases we prove below have $\text{SI}(w) \leq (L^2 - 1)/4$, so the function constructed here does not give a sharp bound. ■

3 Odd length words

3.1 A lower bound for the maximal self-intersection number; proof of Theorem 1.1 (3)

Proof of Theorem 1.1, (3) (The maximum self-intersection number for words of odd length L is at least $(L^2 - 1)/4$). We will show that the words of the form $a(aB)^{\frac{L-1}{2}}$ have self-intersection equal to $(L^2 - 1)/4$. Consider a representative of w as in Figure 5, where $n = \frac{L-1}{2}$. There is an $n \times n$ grid of intersection points in the center, plus the n additional intersections p_2, \dots, p_{2n} , a total of $n^2 + n = (L^2 - 1)/4$. We need to check that none of these intersections spans a bigon (this is the only way [9] that an intersection can be deformed away).

With notation from Figure 5, the only vertices that could be part of a bigon are those from which two segments exit along the same edge, i.e. p_2, p_4, \dots, p_{2n} . If we follow the segments from p_2 through edge A they lead to 1 on edge A and $2n + 1$ on edge b , so no bigon there; the segments from p_4 through edge A lead to 3, $2n + 1$ on edge b , to 2, $2n$ on edge A and then to 1 on edge A and $2n - 1$ on edge b , so no bigon; etc. Finally the segments from p_{2n} through edge A lead to $2n - 1, 2n + 1$ on edge b and eventually to 1 on edge A and 3 on edge b : no bigon. ■

3.2 Preliminaries for upper-bound calculation

In the analysis of self-intersections of odd length words the exact relation between L (the length of a word) and n (its number of $\alpha\beta$ -blocks) becomes more important.

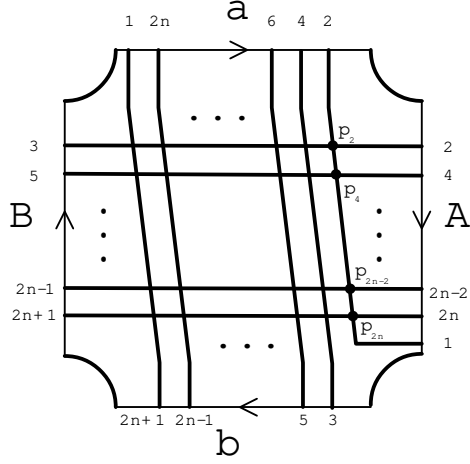


Figure 5: The curve $a(aB)^n$ represented in the fundamental domain for the doubly punctured disc.

Proposition 3.1. *If a word w has length L and n $\alpha\beta$ -blocks, with $L \geq 3n$, then $\text{SI}(w) \leq \frac{1}{4}(L^2 - 1)$. Note that by Theorem 1.1 (3), this estimate is sharp.*

Proof. As established in the previous section (equation 1) $\text{SI}(w) \leq nL - n^2 + n - 1$.

The inequality $nL - n^2 + n - 1 \leq \frac{1}{4}(L^2 - 1)$ is equivalent to $L^2 - 4nL + 4n^2 - 4n + 3 \geq 0$. As a function of L this expression has two roots: $2n \pm \sqrt{4n - 3}$; as soon as L is past the positive root, the inequality is satisfied.

If $n \geq 3$, then $L \geq 3n$ implies $L \geq 2n + \sqrt{4n - 3}$.

If $n = 2$ our inequality $\text{SI}(w) \leq nL - n^2 + n - 1$ translates to $\text{SI}(w) \leq 2L - 3$ which is less than $\frac{1}{4}(L^2 - 1)$ always.

If $n = 1$ our inequality becomes $\text{SI}(w) \leq L - 1$, which is less than $\frac{1}{4}(L^2 - 1)$ as soon as $L \geq 3$. The other only possibility is $L = 2$, an even length. \blacksquare

3.3 The cases: n prime or n a power of 2; proof of Theorem 1.4

Other results for odd-length words require a more detailed analysis of the functions $f \circ r^i(a_1, \dots, a_n, b_1, \dots, b_n)$, keeping the notation of the previous section.

The proof of the following results is straightforward.

Lemma 3.2. *For a fixed $(a_1, \dots, a_n, b_1, \dots, b_n)$, set*

$$s_a = a_1 + \dots + a_n,$$

$$s_b = b_1 + \dots + b_n,$$

$$t_i = f \circ r^i(a_1, \dots, a_n, b_1, \dots, b_n).$$

Then

- (i) $t_{i+1} - t_i = 2n(a_i - b_i) - 2(s_a - s_b)$
- (ii) $t_0 - t_{n-1} = 2n(a_n - b_n) - 2(s_a - s_b)$.
- (iii) $t_{i+j} - t_i = 2n(a_i + \cdots + a_{i+j-1} - b_i - \cdots - b_{i+j-1}) - 2j(s_a - s_b)$.

In particular, if $t_i = t_{i+r}$, for some $r > 0$, then

$$n(a_1 - b_1 + a_2 - b_2 + \cdots + a_{i+r-1} - b_{i+r-1}) = r(s_a - s_b).$$

Lemma 3.3. *If n is prime and $L < 3n$, then all the numbers t_0, \dots, t_{n-1} are different.*

Proof. By Lemma 3.2, if $t_i = t_{i+r}$, for some $r > 0$, then n must divide r or $s_a - s_b$. We will show each is impossible. The first cannot happen because $r < n$. As for the second, observe that $s_a \geq n$ and $s_b \geq n$, and that their sum is $L < 3n$; so $s_a - s_b = s_a + s_b - 2s_b < 3n - 2n = n$. So n cannot divide $s_a - s_b$ either. ■

Lemma 3.4. *If n is a power of 2 and L is odd, then all the numbers t_0, \dots, t_{n-1} are different.*

Proof. Arguing as in Lemma 3.3: in this case, since $r < n$ it cannot be a multiple of n , so $s_a - s_b$ must be even. But $s_a - s_b$ is congruent mod 2 to $s_a + s_b = L$, which is odd. ■

Proposition 3.5. *If a word w of odd length L has a number of $\alpha\beta$ -blocks which is prime or a power of two then $\text{SI}(w) \leq (L^2 - 2)/4$.*

Proof. Let n be the number of $\alpha\beta$ -blocks in w . By Lemmas 3.3 and 3.4 the numbers t_0, \dots, t_{n-1} are all different; in fact (Lemma 3.2) their differences are all even, so any two of them must be at least 2 units apart. It follows that

$$\sum_{i=0}^{n-1} t_i \geq \min t_i + (\min t_i + 2) + \cdots + (\min t_i + 2n - 2) = n \min t_i + n(n - 1)$$

so their average, which we calculated in the proof of Proposition 2.4 to be nL , is greater than or equal to $\min t_i + n - 1$, and so (using equation 1)

$$\text{SI}(w) \leq \min t_i - n^2 + n - 1 \leq nL - n^2 = n(L - n) \leq L^2/4;$$

since L is odd and $\text{SI}(w)$ is an integer, this means

$$\text{SI}(w) \leq (L^2 - 1)/4.$$

■

Propositions 3.1 and 3.5 prove Theorem 1.4.

4 Lower bounds; proof of Theorem 1.5

Definition 4.1. A word in the generators of a surface group and their inverses is *positive* if no generator occurs along with its inverse. Note that a positive word is automatically cyclically reduced. ■

Notation 4.2. If w is a word in the alphabet $\{a, A, b, B\}$, we denote by $\alpha(w)$ (resp. $\beta(w)$) the total number of occurrences of a and A (resp. b and B).

Proposition 4.3. *For any reduced cyclic word w in the alphabet $\{a, A, b, B\}$ there is a positive cyclic word w' of the same length with $\alpha(w') = \alpha(w)$, $\beta(w') = \beta(w)$ and $\text{SI}(w') \leq \text{SI}(w)$.*

Proof. We show how to change w into a word written with only a and b while controlling the self-intersection number. If all the letters in w are capitals, take $w' = w^{-1}$. Otherwise, look in w for a maximal (cyclically) connected string of (one or more) capital letters. The letters at the ends of this string must be one of the pairs (A, A) , (A, B) , (B, A) , (B, B) . In the case (B, B) (the other three cases admit a similar analysis), focus on that string and write

$$w = \langle xa^{a_1} B^{b_1} A^{a_2} B^{b_2} \dots A^{a_i} B^{b_i} a^{a_{i+1}} \rangle$$

where x stands for the rest of the word.

Consider a representative of w with minimal self-intersection. In this representative consider the arcs corresponding to the segments aB (joining the last a of the a^{a_1} -block to the first B of B^{b_1}) and Ba (joining the last B in B^{b_i} to the first a in $a^{a_{i+1}}$). These two arcs intersect in a point p . Perform surgery around p in the following way: remove these two segments, and replace them with an ab and a ba respectively, using the same endpoints. This surgery links the arc $a^{a_{i+1}}xa^{a_1}$ to the arc $B^{b_1}A^{a_2}B^{b_2} \dots A^{a_i}B^{b_i}$ traversed in the opposite direction, i.e. gives a curve corresponding to the word

$$w' = \langle a^{a_{i+1}}xa^{a_1}(B^{b_1}A^{a_2}B^{b_2} \dots A^{a_i}B^{b_i})^{-1} \rangle.$$

This word has the same α and β values as w , has lost at least one self-intersection, and has strictly fewer upper-case letters than w . The process may be repeated until all upper-case letters have been eliminated. ■

Proposition 4.4. *In any surface S with boundary, Let w be a cyclically reduced word in the generators of $\pi_1 S$ which does not admit a simple representative curve. Then a linear word \mathbf{w} representing w (notation from Section 2) can be written as the concatenation $\mathbf{w} = \mathbf{u} \cdot \mathbf{v}$ of two linear words, in such a way that the associated cyclic words satisfy $\text{SI}(u) + \text{SI}(v) + 1 \leq \text{SI}(w)$. (Note that u and v are not necessarily cyclically reduced).*

Proof. Consider a minimal representative of w drawn in the fundamental domain. It must have self-intersections; let p be one of them. Let $\mathbf{w} = x_1x_2 \dots x_L$, (where $x_i \in$

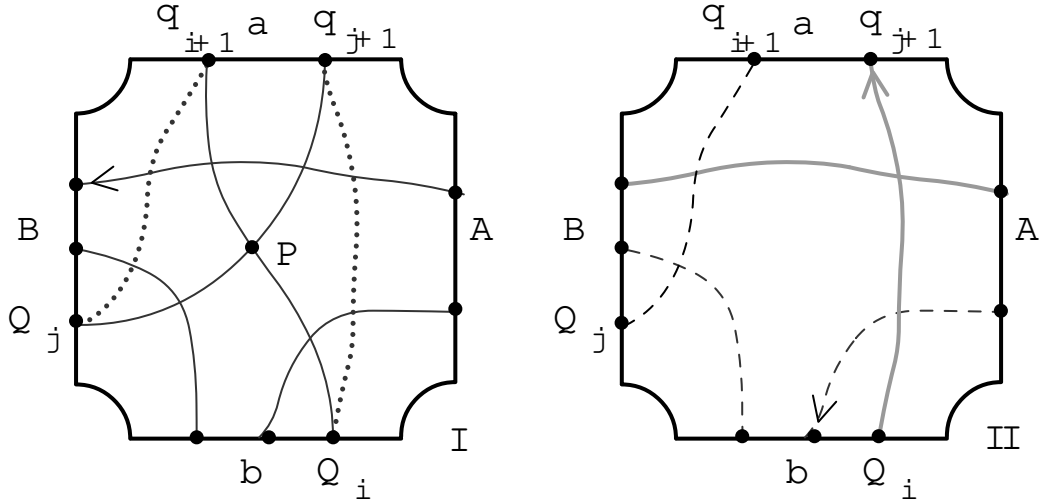


Figure 6: Splitting w as $u \cdot v$ does not add any new intersections, while the intersection corresponding to p is lost. This figure shows $w = Babba$ (I) yielding $u = aB$ and $v = bba$ (II).

$\{a, A, b, B\}$), be a linear representative for w , and suppose that $x_i x_{i+1}$ and $x_j x_{j+1}$, with $i < j$, are the two segments intersecting at p , (see Figure 6, where $x_i x_{i+1} = Ba$ and $x_j x_{j+1} = ba$). Set $u = x_{j+1} \dots x_L x_1 x_2 \dots x_i$ and $v = x_{i+1} \dots x_j$. (In case $i + 1 = j$, v is a single-letter word). The cyclic words u and v together contain all the segments of w , except that $x_i x_{i+1}$ and $x_j x_{j+1}$ have been replaced by $x_i x_{j+1}$ and $x_j x_{i+1}$.

Furthermore, there is a one-to-one correspondence between the intersection points on $x_i x_{j+1} \cup x_j x_{i+1}$ and some subset of the intersection points on $x_i x_{i+1} \cup x_j x_{j+1}$. In fact, labeling the endpoints of the segment corresponding to $x_i x_{i+1}$ (resp. $x_j x_{j+1}$) as Q_i and q_{i+1} (resp. Q_j and q_{j+1}), as in Figure 6, observe that the segment corresponding to $x_i x_{j+1}$ and the broken arc $Q_i p q_{j+1}$ have the same endpoints, so any segment intersecting the first must intersect the second and therefore intersect part of $x_i x_{i+1} \cup x_j x_{j+1}$; similarly for $x_j x_{i+1}$ and $Q_j p q_{i+1}$ (compare Figure 6). Therefore the change from w to $u \cup v$ does not add any new intersections, while the intersection corresponding to p is lost. Hence $\text{SI}(u) + \text{SI}(v) + 1 \leq \text{SI}(w)$. \blacksquare

The next lemma is needed in the proof of Proposition 4.6.

Lemma 4.5. *In the doubly punctured plane P , if a reduced, non-empty word has a simple representative curve, then that curve is parallel to a boundary component. Thus with the notation of Figure 1 the only such words are a, b, ab, A, B and AB .*

Proof. Let γ be a simple, essential curve in P . Since P is planar, $P \setminus \gamma$ has two connected components, P_1 and P_2 . Since γ is essential, neither P_1 nor P_2 is contractible, hence their Euler characteristics satisfy $\chi(P_1) \leq 0$ and $\chi(P_2) \leq 0$; since $\chi(P) = -1$ and $\chi(P) = \chi(P_1) + \chi(P_2)$ it follows that either $\chi(P_1) = 0$ or $\chi(P_2) = 0$. Hence, one of the

two connected components is an annulus, which implies that γ is parallel to a boundary component, as desired. ■

Proposition 4.6. *If w is a positive cyclic word representing a free homotopy class in the doubly punctured plane then $SI(w) \geq \alpha(w) - 1$ and $SI(w) \geq \beta(w) - 1$.*

Proof. By Lemma 4.5 the only words corresponding to simple curves are a, b, ab and their inverses; for these, the statement holds. In particular it holds for all words of length one. Suppose w is any other positive word; it has length L strictly greater than 1. We may suppose by induction that the statement holds for all words of length less than L . By Proposition 4.4, since the curve associated to w is non-simple, the word w has a linear representative \mathbf{w} which can be split as $\mathbf{u} \cdot \mathbf{v}$ so that the associated cyclic words satisfy $SI(w) \geq SI(u) + SI(v) + 1$. Note that u and v have length strictly less than L ; furthermore since w is positive, so are u and v . Therefore by the induction hypothesis $SI(u) + SI(v) + 1 \geq \alpha(u) - 1 + \alpha(v) - 1 + 1$, and so $SI(w) \geq \alpha(u) + \alpha(v) - 1 = \alpha(w) - 1$. The β inequality is proved in the same way. ■

Proof of Theorem 1.5 By Proposition 4.3 there is a positive word w' of length L such that $\alpha(w') = \alpha(w)$, $\beta(w') = \beta(w)$ and $SI(w) \geq SI(w')$. Then Proposition 4.6 yields $SI(w') \geq \max\{\alpha(w), \beta(w)\} - 1$. Since $\alpha(w) + \beta(w) = L$ it follows that $SI(w) \geq L/2 - 1$ if L is even and $SI(w) \geq (L + 1)/2 - 1 = (L - 1)/2$ if L is odd. ■

References

- [1] <http://www.math.sunysb.edu/~moira/applets/chrisApplet.html>
- [2] A. Basmajian, The stable neighborhood theorem and lengths of closed geodesics, *Proc. Amer. Math. Soc.* **119**(1) (1993) 217-224.
- [3] J. Birman and C. Series, An algorithm for simple curves on surfaces, *J. London Math. Soc.* (2), **29** (1984), 331-342.
- [4] M. Bridson and A. Haefliger, *Metric Spaces of Non-Positive Curvature*, Springer-Verlag, Berlin Heidelberg New York, 1991.
- [5] P. Buser, *Geometry and Spectra of Compact Riemann Surfaces*, Birkhauser Boston, 1992.
- [6] <http://www.math.sunysb.edu/~moira/applets/intersectionApplet.html>
- [7] M. Chas and A. Phillips, Self-intersection of curves on the punctured torus, to appear in *Experimental Mathematics*.
- [8] M. Cohen and M. Lustig, Paths of geodesics and geometric intersection numbers I, *Combinatorial Group Theory and Topology*, Alta, Utah, 1984, *Ann. of Math. Studies* **111**, Princeton Univ. Press, Princeton, (1987), 479-500.

- [9] J. Hass and P. Scott, Intersections of curves on surfaces, *Israel J. Math.* **51** (1985) 90-120.
- [10] S. Lalley, Self-intersections of closed geodesics on a negatively curved surface: statistical regularities, *Convergence in ergodic theory and probability* (Columbus, OH, 1993), Ohio State Univ. Math. Res. Inst. Publ., 5, de Gruyter, Berlin, 1996, pp. 263-272. MR1412610 (97h:58103)
- [11] G. Margulis, Applications of ergodic theory to the investigation of manifolds of negative curvature, *Funct. Anal. and Appl.* **3** (1983) 573-591.
- [12] M. Mirzakhani, Growth of the number of simple closed geodesics on a hyperbolic surface, *Ann. of Math.* **168** (2008) 97-125.
- [13] I. Rivin, Simple curves on surfaces, *Geom. Dedicata* **87** (2001), 345-360.
- [14] I. Rivin, Geodesics with one self-intersection and other stories, [arXiv:0901.2543](https://arxiv.org/abs/0901.2543)[math.GT]

DEPARTMENT OF MATHEMATICS, STONY BROOK UNIVERSITY, STONY BROOK NY
11794

E-mail: moira@math.sunysb.edu, tony@math.sunysb.edu