Almost simple geodesics on the triply–punctured sphere

Moira Chas, Curtis T. McMullen and Anthony Phillips

5 March 2017

Abstract

In this paper we study closed hyperbolic geodesics γ on the triply–punctured sphere $M = \widehat{\mathbb{C}} - \{0, 1, \infty\}$ that are *almost simple*, in the sense that the difference $\delta = I(\gamma) - L(\gamma)$ between the self–intersection number of γ and its combinatorial (word) length is fixed. We show that for each fixed δ, the number of almost simple geodesics with $L(\gamma) = L$ is given by a quadratic polynomial $p_\delta(L)$, provided $L \geq \delta + 4$.

Contents

Research supported in part by the NSF. MSC 30F60, 57M05. Typeset 2019-01-26 17:48.

1 Introduction

This paper describes a new phenomenon in the enumeration of closed curves on surfaces.

There are no simple closed geodesics on the triply–punctured sphere. That is, the geometric self–intersection number $I(\gamma)$ of every closed hyperbolic geodesic γ on the Riemann surface

$$M = \widehat{\mathbb{C}} - \{0, 1, \infty\}$$

(endowed with its complete conformal metric of constant curvature -1) satisfies $I(\gamma) > 0$.

In the absence of simple loops, one can aim instead to classify and enumerate those geodesics on M that are *almost simple*, in the sense that $I(\gamma)$ is small compared to the *combinatorial length* $L(\gamma)$. For our purposes, it is convenient to define $L(\gamma)$ to be the number of times that γ passes through the upper halfplane; equivalently, $2L(\gamma)$ is the number of times that γ crosses the real line $\mathbb{R} \subset \widehat{\mathbb{C}}$ (for an example, see Figure 1). Our first result (§4) relates these two quantities.

Theorem 1.1 *For any closed geodesic $\gamma \subset M$, we have $I(\gamma) \geq L(\gamma) - 1$.*

The *defect*

$$\delta(\gamma) = I(\gamma) - L(\gamma) \geq -1$$

is thus a natural measurement of the failure of γ to be simple. For a typical long geodesic, $\delta(\gamma)$ is on the order of $L(\gamma)^2$ (see e.g. [CP2]). We will be interested in the opposite regime, where $\delta(\gamma) = O(1)$. More precisely, for each fixed δ we wish to study the function

$$N_\delta(L) = |\{\gamma \subset M \; : \; L(\gamma) = L \quad \text{and} \quad \delta(\gamma) = \delta\}|.$$

Here we have identified γ with a subset of M, so $N_\delta(L)$ is a count of the number of *unoriented, primitive* geodesics of length L and defect δ. Table 2 gives the value of $N_\delta(L)$ for small δ and L.

Our main result is:

Theorem 1.2 (Quadratic enumeration) *For each $\delta \geq -1$, there exists a quadratic polynomial $p_\delta(L)$ such that $N_\delta(L) = p_\delta(L)$ for all $L \geq \delta + 4$.*

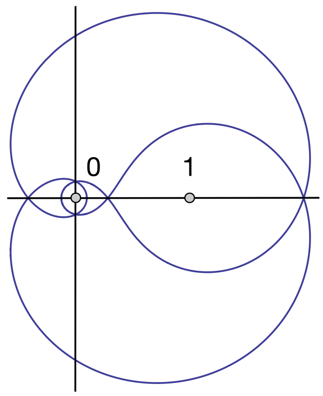

Figure 1. A closed hyperbolic geodesic with $L(\gamma) = 4$ and $I(\gamma) = 5$.

We emphasize that Theorem 1.2 concerns the exact value of $N_\delta(L)$, not just its asymptotic behavior. The polynomials $p_\delta(L)$ for $\delta \le 11$ can be found by examining the columns of Table 2; for example, we have:

$$\begin{aligned} p_{-1}(L) &= 3L^2 - 9L + 9, \\ p_0(L) &= 4L^2 - 24L + 38, \\ p_1(L) &= 30L^2 - 240L + 486, \quad \text{and} \\ p_{11}(L) &= 16608L^2 - 363900L + 2030832. \end{aligned}$$

The statements of Theorems 1.1 and 1.2 were first suggested by the experimental data in this table.

Question. Is it true that $N_\delta(L) = 0$ if and only if $\delta > (L-3)(L-1)/3$?

Pairs of pants. Although stated in terms of geodesics, Theorems 1.1 and 1.2 can be regarded as topological results about closed loops on M (or equivalently, on a pair of pants). Indeed, every essential, nonperipheral closed loop on M is homotopic to a unique geodesic γ, and $I(\gamma)$ is simply the minimum number of (transverse) self-intersections among all representatives of that homotopy class. The combinatorial length $L(\gamma)$ can also be described topologically, in terms of generators for $\pi_1(M)$ (see §2). For more details on the geometric intersection number, see e.g. [FLP], [Bon].

Decorations. The mechanism behind Theorem 1.2 is illustrated in Figure 3. The trefoil at the left in the figure shows a geodesic with $I(\gamma) = L(\gamma) = 3$, and hence $\delta(\gamma) = 0$. For each $(i, j, k) \in \mathbb{Z}^3$ we can decorate γ by adding multiple loops around each of the three punctures of M to obtain the homotopy class of another geodesic with $\delta(\gamma_{ijk}) = 0$; here the signs of i, j

	$\delta = -1$	0	1	2	3	4	5	6	7	8	9	10	11
$L = 2$	3	0	0	0	0	0	0	0	0	0	0	0	0
3	9	1	0	0	0	0	0	0	0	0	0	0	0
4	21	6	3	0	0	0	0	0	0	0	0	0	0
5	39	18	36	9	0	0	0	0	0	0	0	0	0
6	63	38	126	54	27	18	9	0	0	0	0	0	0
7	93	66	276	156	216	150	135	51	21	6	0	0	0
8	129	102	486	318	666	528	672	438	375	180	78	72	36
9	171	146	756	540	1386	1218	2070	1648	1995	1269	1088	948	660
10	219	198	1086	822	2376	2226	4560	4044	5970	4632	5532	4890	4596
11	273	258	1476	1164	3636	3552	8160	7764	13302	11571	16608	15342	18081
12	333	326	1926	1566	5166	5196	12870	12818	24414	22806	36779	35838	49428
13	399	402	2436	2028	6966	7158	18690	19206	39336	38574	67836	68925	105708
14	471	486	3006	2550	9036	9438	25620	26928	58068	58890	110454	115806	191337
15	549	578	3636	3132	11376	12036	33660	35984	80610	83754	164678	176844	309132
16	633	678	4326	3774	13986	14952	42810	46374	106962	113166	230508	252060	460080
17	723	786	5076	4476	16866	18186	53070	58098	137124	147126	307944	341454	644244

Table 2. The number $N_\delta(L)$ of unoriented, primitive geodesics on $\widehat{\mathbb{C}} - \{0, 1, \infty\}$ of combinatorial length L and self-intersection number $I = L + \delta$.

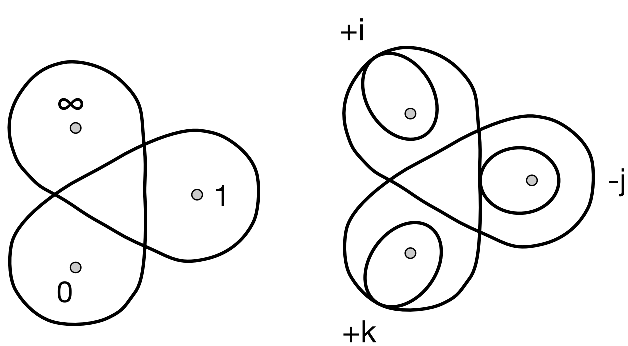

Figure 3. Topology of closed geodesics with $\delta(\gamma) = 0$.

and k indicate if the decorations are to be attached to the inner or outer triangle of γ. The homotopy class of γ_{ijk} is indicated schematically by the train–track shown at the right. It turns out that every geodesic with $\delta = 0$ is obtained in this way, and hence $N_0(L)$ is simply the number of solutions to the equation

$$L(\gamma_{ijk}) = 3 + |i| + |j| + |k| = L.$$

Explicitly, this count is given by

$$N_0(L) = 8\binom{L-4}{2} + 12\binom{L-4}{1} + 6\binom{L-4}{0} + \binom{L-4}{-1}, \qquad (1.1)$$

which agrees with the quadratic polynomial $p_0(L)$ for $L \geq 4$.

Binomial coefficients: conventions. In the statement of this and other results, we adopt the convention that

$$\binom{n}{k} = 0 \text{ if } n \text{ or } k \text{ is negative, except } \binom{-1}{-1} = 1. \qquad (1.2)$$

This convention is chosen so that the equation

$$\binom{n+r-1}{r-1} = \left|\left\{(n_1, \ldots, n_r) \;:\; n_i \geq 0, n_i \in \mathbb{Z}, \sum_1^r n_i = n\right\}\right| \qquad (1.3)$$

is valid for all integers n and all $r \geq 0$. With this convention, the usual expression for $\binom{n}{k}$ as a polynomial in n is valid only for $n \geq 0$. For example, formula (1.1) for $N_0(L)$ agrees with the quadratic polynomial $p_0(L)$ for $L \geq 4$, but for $L = 1, 2, 3$ we have $N_0(L) = 0, 0, 1$ while $p_0(L) = 18, 6, 2$.

Motifs. More generally, to prove Theorem 1.2, we will show that every closed geodesic in M with $\delta(\gamma) = \delta$ is obtained by decorating one of finitely many special geodesics called *motifs*. In terms of these motifs, we obtain a formula for $N_\delta(L)$ as a sum of binomial coefficients, valid for all L (§5):

Theorem 1.3 (Binomial enumeration) *For all integers δ and L, we have*

$$N_\delta(L) = \sum_{\textit{motifs } \gamma \textit{ with } \delta(\gamma) = \delta} \binom{L - L(\gamma) + \rho(\gamma) - 1}{\rho(\gamma) - 1}.$$

Here the *rank* of a motif, $0 \leq \rho(\gamma) \leq 3$, indicates how many decorations it admits.

For example, there are 27 motifs with $\delta(\gamma) = 0$; they correspond, in terms of Figure 3, to the geodesics γ_{ijk} with $i, j, k \in \{-1, 0, 1\}$. Three of

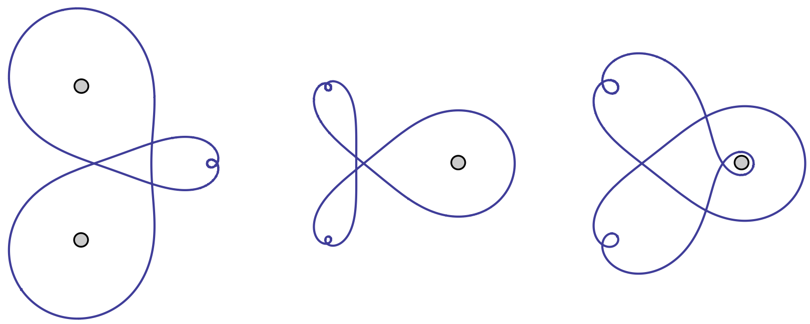

Figure 4. The closed geodesics γ_{010}, γ_{101} and $\gamma_{1,-1,1}$.

these geodesics are shown in Figure 4. Grouping them together according their ranks, given by $\rho(\gamma_{ijk}) = |i|+|j|+|k|$, we obtain the 4 terms in equation (1.1) for $N_0(L)$. For more details on this example and others, see §5.

Lengths of motifs. In §6 we show that the length of any motif satisfies the bound:

$$\delta(\gamma) + \rho(\gamma) + 3 \geq L(\gamma). \tag{1.4}$$

This crucial bound has two important consequences. First, it implies that Theorem 1.3 expresses $N_\delta(L)$ as a *finite* sum of binomial coefficients, since $6+\delta \geq L(\gamma)$; and second, it shows that each binomial coefficient agrees with a polynomial in L of degree $\rho(\gamma)-1 \leq 2$ in the range $L \geq \delta+4$. Thus $N_\delta(L)$ itself is a quadratic polynomial in L, proving Theorem 1.2.

We emphasize that Theorem 1.3 gives a formula for $N_\delta(L)$ that is valid for all L, not just $L \geq \delta + 4$. Moreover, motifs have a simple description in terms of combinatorial group theory (see §5). The main subtlety in evaluating this formula comes from the condition $\delta(\gamma) = \delta$, which requires the computation of the self–intersection number of γ.

Question. What is the behavior of the leading coefficient of $p_\delta(L)$? (This coefficient is one–half the number of motifs with defect δ and full rank.)

Spheres with more punctures and surfaces of higher genus. The theory of almost simple geodesics is most lucid in the case of the triply–punctured sphere, because there are no *simple* geodesics to complicate the analysis.

Much of this paper, however, generalizes in a straightforward way to the case where M is an n–times punctured sphere, $n \geq 4$; for example, Theorem 1.3 remains valid in this setting. The crucial difference is that for $n \geq 4$, there are infinitely many motifs with $\delta(\gamma) = \delta$, so the formula for $N_\delta(L)$ in Theorem 1.3 becomes an infinite sum of binomial coefficients. Consequently $N_\delta(L)$ cannot be given by a polynomial in L when $n \geq 4$. Nevertheless, we have an exact formula for $N_\delta(L)$ and the study of closed loops on M can be reduced to the study of motifs.

It would also be interesting to extend the study of almost simple geodesics to punctured surfaces of higher genus. For example, the collection of geodesics that become trivial when the punctures are filled in may admit a similar analysis.

Perspectives on M. The proof of Theorem 1.2 pivots on two crucial shifts in perspective on the geometry and combinatorial group theory of the triply–punctured sphere M.

The first shift is to express words in $G = \pi_1(M, p)$ in terms of generators, not for G, but for the reflection group

$$\widetilde{G} = \langle x, y, z \,:\, x^2 = y^2 = z^2 = \mathrm{id} \rangle$$

which contains G with index two. The advantage of these generators is that they do not privilege the upper or lower halfplane, and that they allow one to give combinatorial meaning to a geodesic which takes a half–integral number of turns around a cusp.

The second shift is to replace the standard hyperbolic metric on M with a complete hyperbolic metric of infinite volume, turning the convex core of M into a symmetric pair of pants with long boundary components. In this new metric, the location of self–intersections of geodesics is changed in an advantageous way, even though the total number of self–intersections remains the same. More precisely, we obtain a direct relationship between the *depth* of a geodesic excursion into one of the ends of M, and the *length* of a run of alternating letters in the corresponding word $w \in \widetilde{G}$.

Outline of the paper. The interplay of hyperbolic geometry and combinatorial group theory just described is used in §3 to bound the change in $I(\gamma)$ when a loop around a cusp is added or removed. A quick proof of Theorem 1.1 follows in §4. The theory of motifs is discussed in §5, leading to the proof of Theorem 1.3, and the proof of Theorem 1.2 is completed in §6.

In the Appendix we give a formula for $I(\gamma)$ in terms of combinatorial group theory, suitable for computing the entries in Table 2. We also establish the lower bound $I(\gamma) \geq L(\gamma)^2/6$ for a class of loops with controlled excursions into the cusps of M. Both results play an important role in §6, where we prove the inequality (1.4)

Notes and references. The asymptotic growth of the number of simple closed geodesics on a hyperbolic surface is studied in [Mir1] and [Er]; see also [Mir2], [ES] and [EPS] for the case of closed geodesics with a fixed self–intersection number. Variants of Table 2, not restricted to primitive geodesics, appear in [CP1] and [CP2]. The statistical distribution of self–intersection numbers is studied in [CL].

For basic background on curves, surfaces and intersection numbers, see e.g. [FLP], [Bon] and [St]; more on intersections numbers can be found in [Re], [CoL], [HS], [DL] and the references therein.

2 Background and notation

In this section we make explicit the relationship between closed geodesics and combinatorial group theory.

Group theory. We will work in the reflection group

$$\widetilde{G} = \langle x, y, z \,:\, x^2 = y^2 = z^2 = \mathrm{id}\rangle,$$

and in the index two subgroup G generated by $(a, b, c) = (xy, yz, zx)$; it has the presentation

$$G = \langle a, b, c \,:\, abc = \mathrm{id}\rangle.$$

A word w in the generators x, y, z of $\widetilde{G}$ is *reduced* if consecutive letters of w are distinct. Every element $w \in \widetilde{G}$ is represented by a unique reduced word, whose length will be denoted by $\ell(w)$. We say w is *cyclically reduced* if its first and last letters are also distinct. The shortest words in a given conjugacy class are cyclically reduced.

We have $w \in G \iff \ell(w)$ is even. We say $w \in G$ is *primitive* unless $w = g^n$ for some $n > 1$ and $g \in G$.

We say two cyclically reduced words $w_1, w_2 \in G$ are *equivalent*, if w_1 is conjugate to $w_2^{\pm 1}$ in G. In terms of words, if $w_1 = g_1 \cdots g_{2n}$, then

$$w_2 = (g_{2i+1} \cdots g_{2n} g_1 \cdots g_{2i})^{\pm 1}$$

for some i. Let

$$\mathcal{G} = \{w \in G \,:\, w \text{ is primitive, cyclically reduced and } \ell(w) \geq 4\} \,/ \sim .$$

Each element of $\mathcal{G}$ is an equivalence class $[\omega]$ of cyclically reduced words. The length condition insures that w involves all three generators x, y, z of $\widetilde{G}$.

Hyperbolic geodesics. We now return to the triply–punctured sphere

$$M = \widehat{\mathbb{C}} - \{0, 1, \infty\}.$$

The Riemann surface M can be presented as the quotient of the upper halfplane $\mathbb{H} = \{z \in \mathbb{C} : \operatorname{Im}(z) > 0\}$ by the group $\Gamma(2) \subset \mathrm{SL}_2(\mathbb{Z})$ consisting of matrices with $A \equiv I \bmod 2$, acting by Möbius transformations. It carries a unique complete, conformal *hyperbolic metric* of constant curvature -1, inherited from the metric $|dz|/\operatorname{Im}(z)$ on $\mathbb{H}$. Geodesics on M are covered by semicircles perpendicular to the boundary in $\mathbb{H}$.

Choosing a basepoint $p \in M$ with $\operatorname{Im}(p) > 0$, we have an isomorphism

$$\phi : \pi_1(M, p) \cong G.$$

To define this map unambiguously, we label the components $(-\infty, 0)$, $(0, 1)$ and $(1, \infty)$ of $\mathbb{R} \cap M$ by X, Y and Z respectively. Given a smooth loop $\gamma : [0, 1] \to M$ transverse to the real axis, with $\gamma(0) = \gamma(1) = p$, let $0 < t_1 < \cdots < t_{2n} < 1$ denote the parameters such that $\gamma(t_i) \in \mathbb{R}$. To each such crossing we associate one of the generators of $\widetilde{G}$, namely $g_i = x$ if $\gamma(t_i) \in X$, $g_i = y$ if $\gamma(t_i) \in Y$, and $g_i = z$ if $t_i \in Z$. We then define

$$\phi([\gamma]) = g_1 \cdots g_{2n}.$$

It is readily verified that ϕ depends only on the homotopy class of γ and gives an isomorphism to G.

This isomorphism determines a natural bijection

$$\mathcal{G} \leftrightarrow \{\text{closed hyperbolic geodesics } \gamma \subset M\}. \tag{2.1}$$

This map sends a conjugacy class in $\pi_1(M, p)$ to its geodesic representative. Conversely, if we traverse a closed geodesic $\gamma \subset M$ starting at a point in the upper halfplane, and write down the corresponding generator of G each time γ crosses the real axis, we obtain a cyclically reduced word $w \in G$ representing the corresponding class $[w] \in \mathcal{G}$. (The elements of $\pi_1(M, p)$ with $\ell(w) \leq 2$ are excluded from $\mathcal{G}$ because they are peripheral or trivial.)

Self–intersection numbers of words. Using the bijection (2.1), we can regard length and self–intersection number as functions of cyclically reduced words in G, defined by $L(w) = L(\gamma_w)$ and $I(w) = I(\gamma_w)$.

Clearly $L(w) = \ell(w)/2$. A combinatorial algorithm for computing $I(w)$ directly in terms of the word w is described in the Appendix.

3 Surgery

In this section we use hyperbolic geometry to prove some purely combinatorial statements about the behavior of intersection numbers under elementary modifications of a group element $w \in G$.

Overview. The main results of this section are Theorems 3.1, 3.2 and 3.3 below. They are used as tools in all the sections that follow.

Although the statements of these theorems are combinatorial, their proofs are geometric. To carry them out, we first open up the cusps of M to obtain a hyperbolic pair of pants $M(\lambda)$, whose convex core is bounded by three geodesics (cuffs) of equal length.

Each cuff of $M(\lambda)$ induces an annular covering space $\widetilde{M}(\lambda) \to M(\lambda)$ (Figure 6). In this covering space, the winding number $N(\gamma)$ of a geodesic lifted from $M(\lambda) \cong M$ can be related to runs of alternating letters in the corresponding word $w \in \pi_1(M)$ (see Figure 7). On the other hand, the larger the winding number, the closer γ comes to one of the cuffs of $M(\lambda)$ (see equation (3.1) and Figure 5). The part of a geodesic closest to the cuff can be decorated without adding any unexpected new self–intersections, and the desired bounds on $I(w)$ follow.

Runs. We turn to the statements of the main theorems of this section.

Consider a combinatorial closed geodesic $[w] \in \mathcal{G}$, represented by a cyclically reduced word w. Since w is primitive and $\ell(w) \geq 4$, all three generators x, y, z occur as letters of w.

A *run* r of w is a maximal sequence of (cyclically) consecutive letters of w in which at most two generators appear. (The location of r in w is part of the information determining a run.) Clearly $2 \leq \ell(r) < \ell(w)$.

If r only involves the generators x, y, then we say r is a run of *type* xy. Any run is of type xy, yz or zx.

Expansion. Given a run r of w, we can *expand* r by repeating its first two letters to obtain a word r^+ with $\ell(r^+) = \ell(r) + 2$; and then expand w by replacing r with r^+. The resulting word $w^+[r]$ is still cyclically reduced, because the first and last letters of r are the same as those of r^+. In particular, we have:

$$L(w^+[r]) = L(w) + 1.$$

For example, if $w = xyzy$, and r is the initial run xy, then $w^+[r] = xyxyzy$.

Contraction. Similarly, given any run r of w with $\ell(r) \geq 3$, we can *contract* r to r^- by removing its first two letters, and then contract w by replacing r

with r^-. The resulting word $w^-[r]$ is still reduced, and hence

$$L(w^-[r]) = L(w) - 1.$$

For example, if $w = xzxyxy$, and r is the initial run xzx, then $w^-[r] = xyxy$. This example shows that primitivity need not be preserved by contraction.

Behavior of $I(w)$. The main results of this section control the behavior of the self–intersection number $I(w)$ under expansion and contraction. Let us say a run is *exceptional* if

$$I(w^-[r]) = I(w) - 1.$$

Theorem 3.1 *A cyclically reduced word w with $[w] \in \mathcal{G}$ has at most one exceptional run of each type.*

Theorem 3.2 *Let r be a run of w with $\ell(r) \geq 3$. Then either:*

$$I(w^-[r]) = I(w) - 1,$$

and $\ell(r) \geq \ell(s)$ for every other run s of the same type as r; or

$$I(w^-[r]) \leq I(w) - 3.$$

Theorem 3.3 *If $\ell(r) > \ell(s)$ for every other run s of the same type as r, then $I(w^+[r]) = I(w) + 1$.*

Since contraction decreases the length of a run by two, we have:

Corollary 3.4 *If $\ell(r) > \ell(s) + 2$ for every other run s of the same type as r, then r is exceptional: we have $I(w^-[r]) = I(w) - 1$.*

Hyperbolic geodesics in annuli. The proofs of these results are based on hyperbolic geometry. Fix $0 < r < 1$. We begin by recalling some simple facts about the annulus

$$U(r) = \{z \,:\, r < |z| < 1\},$$

considered as a hyperbolic Riemann surface.

First, by symmetry, its unique closed geodesic is the circle $A \subset U(r)$ defined by $|z| = \sqrt{r}$. If we denote the length A by $\log(a)$, $a > 1$, then the universal covering map $f : \mathbb{H} \to U(r)$ is given by

$$f(t) = \exp(2\pi i \, \log(t)/\log(a)).$$

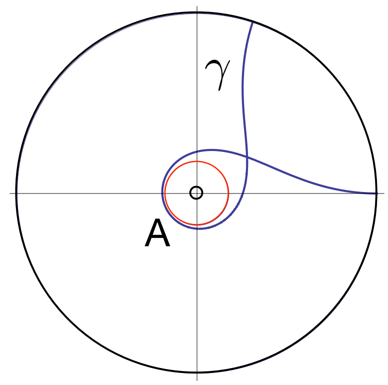

Figure 5. A hyperbolic geodesic γ in an annulus with core geodesic A; its winding number satisfies $N(\gamma) > 1$.

The deck group of $\mathbb{H}/U(r)$ is generated by $z \mapsto az$, and A is the image of the positive imaginary axis under f. The values of a and r are related by $r = f(-1) = \exp(-2\pi^2/\log(a))$.

Winding numbers. Now consider an immersed hyperbolic geodesic $\gamma \subset U(r)$ with both endpoints on the unit circle S^1 (see Figure 5). Any such geodesic in $U(r)$ can be presented as the image of an embedded geodesic $\widetilde{\gamma} \subset \mathbb{H}$ with endpoints $0 < x_1 < x_2$ on the positive real axis. The *winding number* of γ is the unique real number $N(\gamma) > 0$ such that

$$x_2 = a^{N(\gamma)} x_1.$$

It is easily seen that γ winds $N(\gamma)$ times around the annulus $U(r)$, in the sense that the angle $\arg(z)$ increases strictly monotonically by $2\pi N(\gamma)$ as z moves counter-clockwise from one end of γ to the other.

The distance $d(z, A)$ is a strictly convex function on γ; in particular, it assumes its minimum value at a unique point p. The more γ winds around the annulus, the closer it comes to its core geodesic A; that is, the minimum distance satisfies

$$N(\gamma) > N(\gamma') \implies d(\gamma, A) < d(\gamma', A). \tag{3.1}$$

This follows immediately from the fact that the hyperbolic distance from $\widetilde{\gamma}$ to the imaginary axis is a decreasing function of x_2/x_1.

Self–intersections. When the winding number $N(\gamma)$ exceeds 1, the geodesic γ has 1 or more self–intersections. (To see this, just observe that the endpoints (ax_1, ax_2) of $a\widetilde{\gamma}$ are linked with the endpoints (x_1, x_2) of $\widetilde{\gamma}$ when

$x_2 > ax_1$.) In particular, γ has a unique self–intersection point q which is closest to A. (The point q and the point of closest approach p are opposite one another, i.e. $\arg(q) = \arg(-p)$.) The self–intersection point q cuts off an embedded loop $\gamma^0 \subset \gamma$ encircling A, and the region between A and γ^0 is a convex annulus in the hyperbolic metric on $U(r)$.

If we remove γ^0 from γ, we are left with a path $\gamma^- \subset U(r)$ whose geodesic representative has winding number $N(\gamma) - 1$.

Symmetric pairs of pants. To study the triply–punctured sphere $M = \mathbb{H}/\Gamma(2)$, it turns out to be useful to consider instead the symmetric pair of pants $M(\lambda)$ with cuffs of length $4\log\lambda$. (It is well known that there is a unique pair of pants with cuffs of given lengths; see e.g. [IT, §3.1.5].)

Given $\lambda > 1$, a concrete model for the surface $M(\lambda)$ can be constructed as follows. First, consider the hyperbolic geodesics $X, Y \subset \mathbb{H}$ defined by $|t| = \lambda^{-1}$ and $|t| = \lambda$ respectively. Let $Z \subset \mathbb{H}$ be the unique geodesic disjoint from these two, with endpoints on the positive real axis, such that

$$d(X, Y) = d(Y, Z) = d(Z, X).$$

The endpoints of Z are given by α and α^{-1}, where $1 < \alpha < \lambda$. The value of α is uniquely determined by λ and satisfies

$$\alpha(\lambda) \sim 1 + 1/\lambda \tag{3.2}$$

as $\lambda \to \infty$. It can be computed using cross-ratios and the fact that there is a Möbius transformation that cyclically permutes X, Y and Z.

Letting $x, y, z \in \mathrm{Isom}(\mathbb{H})$ denote the reflections through the geodesics X, Y and Z respectively, we obtain a natural isometric action of the group

$$\widetilde{G} = \{x, y, z \,:\, x^2 = y^2 = z^2 = \mathrm{id}\}$$

introduced in §2. The orientation–preserving subgroup $G \subset \widetilde{G}$ is generated by xy, yz and zx, and the quotient space

$$M(\lambda) = \mathbb{H}/G$$

is the desired symmetric pair of pants. A fundamental domain for the action of G on $\mathbb{H}$ is given by the region bounded by Y, Y^* and Z, Z^*, where Y^* and Z^* are the images of Y and Z under reflection through X (see the top of Figure 6).

Intersection numbers. Note that as $\lambda \to 1$, the infinite volume surface $M(\lambda)$ converges geometrically to the finite volume surface $M = \mathbb{H}/\Gamma(2)$, which we can regard as $M(1)$.

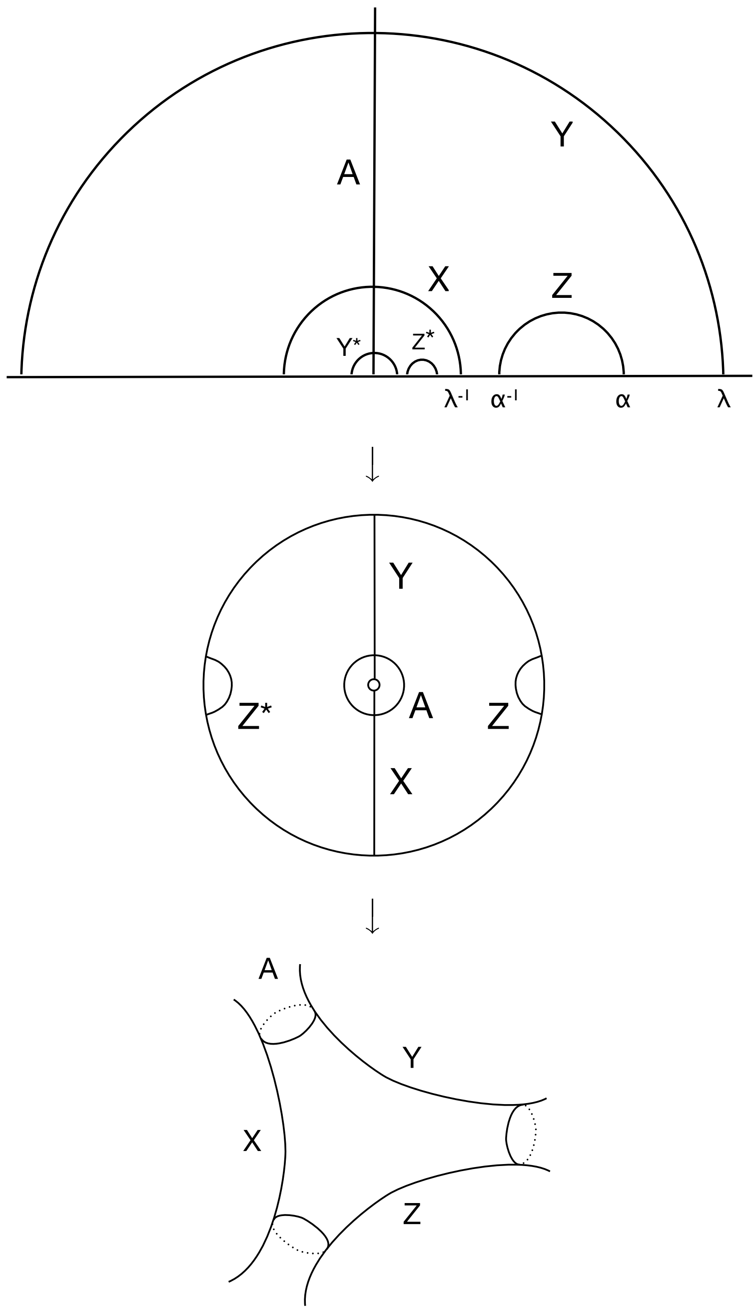

Figure 6. The tower of coverings $\mathbb{H} \to \widetilde{M}(\lambda) \to M(\lambda)$.

The self–intersection number $I(w)$ of a word $w \in G$ can be computed using its geodesic representative on $M(\lambda)$ just as well as on $M(1)$; in fact, geodesics for any hyperbolic metric automatically minimize intersection numbers (see e.g. [FLP, Exposé 3, Theorem 15]). It turns out that the most advantageous geometry for our purposes arises, not when $\lambda = 1$, but when λ is large.

The annular covering space. Note that $xy \in G$ acts on $\mathbb{H}$ by the hyperbolic transformation $t \mapsto a^{-1}t$, where $a = \lambda^4$. This transformation stabilizes the vertical geodesic $A = i\mathbb{R}_+ \subset \mathbb{H}$.

To analyze runs of type xy, we will use the intermediate covering space

$$\widetilde{M}(\lambda) = \mathbb{H}/\langle xy\rangle \cong U(r),$$

where $r = \exp(-2\pi^2/\log a)$ as before. The tower of coverings

$$\mathbb{H} \to \widetilde{M}(\lambda) \to M(\lambda)$$

is shown in Figure 6. The geodesics A, X, Y, Z and Z^* in $\mathbb{H}$ map to geodesics in $\widetilde{M}(\lambda)$ which for simplicity we will continue to denote by the same letters. Note that A descends to the core geodesic of $\widetilde{M}(\lambda)$, while X, Y, Z and Z^* map to embedded geodesics cutting $\widetilde{M}(\lambda)$ into disks.

We now fix any value of λ such that

$$\alpha = \alpha(\lambda) < \lambda^{1/2}.$$

(Such a value exists by (3.2); in fact, any $\lambda \geq 2$ will do.) This choice insures that Z and Z^* each cut off an arc of S^1 of length less than $\pi/2$. As a useful consequence, we can estimate the winding number of a geodesic $\gamma \subset \widetilde{M}(\lambda)$ by combinatorial means.

Lemma 3.5 *Suppose a geodesic $\gamma \subset \widetilde{M}(\lambda) \cong U(r)$ first crosses $Z \cup Z^*$, then crosses $X \cup Y$ a total of ℓ times, and finally crosses $Z \cup Z^*$ again. Then its winding number satisfies*

$$\left| N(\gamma) - \frac{\ell}{2} \right| < \frac{1}{4}.$$

Proof. Suppose for simplicity that γ first crosses Z and winds counterclockwise around $U(r)$. Then γ has a lift to a geodesic $\widetilde{\gamma} \subset \mathbb{H}$ that begins at a point x_1 in the interval bounded by the endpoints (α^{-1}, α) of Z, and ends at a point x_2 in the interval bounded by the endpoints of $\lambda^{2\ell}Z$. It follows that $\lambda^{2\ell}/\alpha^2 \leq x_2/x_1 \leq \lambda^{2\ell}\alpha^2$. Since $N(\gamma) = \log(x_2/x_1)/(4 \log \lambda)$ and $1 < \alpha^2 < \lambda$, the result follows. ∎

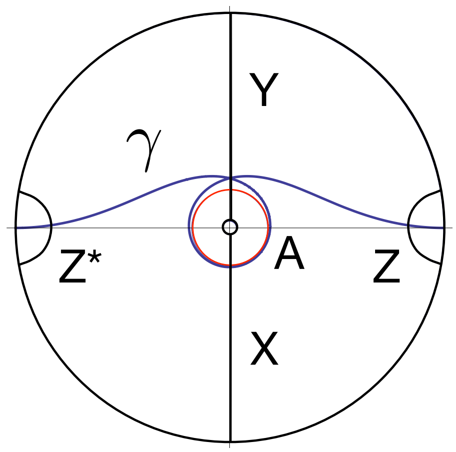

Figure 7. The winding number of a geodesic γ is estimated by half the number of times it crosses $X \cup Y$.

An example of $\gamma \subset \widetilde{M}(\lambda)$ with $\ell = 3$ is shown in Figure 7.

Proof of Theorems 3.1, 3.2 and 3.3. For concreteness we will analyze runs of type xy; the same argument applies to the other two types of runs with only notation modifications.

Given a class $[w] \in \mathcal{G}$, we can choose a representative cyclically reduced word in this class whose first letter is z. Then we have

$$w = zr_1zr_2 \dots zr_n,$$

where the r_i are maximal subwords in which no z appears. The r_i of length two or more are exactly the xy runs of w.

There is a corresponding factorization $\gamma = \gamma_1 * \cdots * \gamma_n$ of the closed geodesic $\gamma \subset M(\lambda)$ representing $[w]$, where the endpoints of each segment γ_i lie on Z and the interior meets $X \cup Y$ a total of $\ell(r_i)$ times.

Let $F \subset \widetilde{M}(\lambda)$ be the region bounded by $Z \cup Z^*$ (see Figure 6). The surface $M(\lambda)$ is obtained from F by gluing its two geodesic boundary components together. Thus we can regard each segment in the factorization of γ as an arc $\gamma_i \subset F$ with endpoints of $Z \cup Z^*$.

Let $\widetilde{\gamma}_i \subset \widetilde{M}(\lambda)$ denote the extension of γ_i to a complete geodesic in the annulus. By Lemma 3.5, the winding number of this geodesic is well–approximated by $\ell(r_i)/2$; that is, we have

$$\left| N(\widetilde{\gamma}_i) - \frac{\ell(r_i)}{2} \right| < 1/4. \tag{3.3}$$

Let I denote the set of indices such that $\ell(r_i) \geq 3$; these are the xy–runs of w where contraction is possible. For each $i \in I$, we have $N(\widetilde{\gamma}_i) > 1$, and hence γ_i has a self–intersection that cuts off an innermost loop γ_i^0 encircling the core geodesic $A \subset \widetilde{M}(\lambda)$. Let γ_i^- denote γ_i with the loop γ_i^0 removed. Then the loop

$$\gamma^- = \gamma_1 * \cdots * \gamma_i^- * \cdots * \gamma_n$$

represents the homotopy class $[w^-[r_i]]$. We now distinguish two cases.

(1) Suppose $\gamma_i^0 \cap \gamma_j \neq \emptyset$ for some $j \neq i$. Then γ has three more self–intersections than γ^-. Indeed, the arc γ_j must enter and exit the annulus bounded by $A \cup \gamma_i^0$, resulting in two self–intersections; and the loop γ_i^0 comes from an intersection of γ_i with itself. Thus:

$$I(w^-[r]) \leq I(w) - 3.$$

In particular, r_i is not exceptional.

(2) Now suppose $\gamma_i^0 \cap \gamma_j = \emptyset$ for all $j \neq i$. Then γ still has one more self–intersection than γ^-, and hence

$$I(w^-[r]) \leq I(w) - 1. \tag{3.4}$$

Moreover, since γ was a geodesic, γ^- admits a regular homotopy to its geodesic representative; thus $I(w^-[r]) = I(w) - 1 \bmod 2$. So in this case, either $I(w^-[r]) = I(w) - 1$ — and r is exceptional; or we have $I(w^-[r]) \leq I(w) - 3$, just as in case (1).

For case (2) to hold, we must have

$$d(\gamma_i^0, A) < d(\gamma_j, A) \tag{3.5}$$

for all $j \neq i$. This shows that (2) holds for at most one value of i, and hence w has at most one exceptional run of type xy (Theorem 3.1). Moreover, (3.5) implies, by relation (3.1), that $N(\widetilde{\gamma}_i) > N(\widetilde{\gamma}_j)$ for all $i \neq j$, and hence $\ell(r_i) \geq \ell(r_j)$ for all $i \neq j$ by equation (3.3). This completes the proof of Theorem 3.2.

Finally we prove Theorem 3.3. Suppose $\ell(r_i) > \ell(r_j)$ for all $j \neq i$. Then, by equations (3.3) and (3.1), we have

$$d(\gamma_i, A) < d(\gamma_j, A)$$

for all $j \neq i$. Let $p \in \gamma_i$ be the point closest to A. By attaching a new loop to γ_i at p that runs once around A, we introduce one new self–intersection and obtain a path

$$\gamma^+ = \gamma_1 * \cdots * \gamma_i^+ * \cdots * \gamma_n$$

representing the class $[w^+[r_i]] \in \mathcal{G}$. This shows that

$$I(w^+[r]) \leq I(w) + 1.$$

On the other hand, if we erase this loop then the self–intersection number goes down by at least one, by Theorem 3.2, and hence equality holds. ■

4 Length and self–intersections

With the surgery bounds in place, it is now straightforward to establish a lower bound on the defect

$$\delta(w) = I(w) - L(w).$$

Theorem 4.1 *For any cyclically reduced word $w \in G$, we have $\delta(w) \geq -1$.*

Multiple loops. As a preliminary, we remark that any primitive, cyclically reduced word $w \in G$ satisfies

$$I(w^n) = n^2 I(w) + n - 1 \quad \text{and} \quad L(w^n) = nL(w). \tag{4.1}$$

The formula for $L(w^n)$ is immediate, and the formula for $I(w^n)$ is well known (see e.g. [dGS, Theorem 6]). In fact, an optimal loop representing $[w^n]$ can be obtained from n parallel copies of an optimal loop for $[w]$ by cyclically braiding the strands to form a single loop. The parallel copies yield $n^2 I(w)$ crossings, and the braiding accounts for $n - 1$ more.

Proof of Theorem 4.1. We first remark that if this bound holds for a primitive, cyclically reduced word $w \in G$, then it also holds for w^n, $n > 1$. Indeed, if we know that $I(w) \geq L(w) - 1$, then by equation (4.1) we also have:

$$I(w^n) = n^2 I(w) + n - 1 \geq n^2 L(w) - n^2 + n - 1 \geq nL(w) - 1 = L(w^n) - 1,$$

and hence $\delta(w^n) \geq -1$ as well.

We can now argue by contradiction. Suppose the bound is false. Let $w \in G$ be a word of minimal length with $\delta(w) < -1$. By the preceding remark, w is primitive. Suppose w has a run r of length three or more, and let $w' = w^-[r]$. Then $L(w') = L(w) - 1$ and, by Theorem 3.2, we have $I(w') \leq I(w) - 1$; hence $\delta(w') \leq \delta(w) < -1$, contradicting the fact that w has minimal length.

Thus w has no run of length 3. But there are only two types of words in G with this property: one is $w = xy$, and the other is $w = xyzxyz$. The first has $\delta(w) = -1$, and the second has $\delta(w) = 0$. ■

Theorem 1.1 follows immediately, by the discussion in §2.

5 Motifs

In this section we will define the set of motifs $\mathcal{M} \subset \mathcal{G}$ and prove Theorem 1.3, which we restate as:

Theorem 5.1 *The number of closed geodesics of length L and defect δ is given by:*

$$N_\delta(L) = \sum_{[w]\in\mathcal{M}\,:\,\delta(w)=\delta} \binom{L - L(w) + \rho(w) - 1}{\rho(w) - 1}.$$

Rank and descendants. Let w be a cyclically reduced word with $[w] \in \mathcal{G}$. Let us say a run r of w is *distinguished* if $\ell(r) > \ell(s)$ for every other run s of the same type as r.

Proposition 5.2 *For any distinguished run r of w, we have $\delta(w^+[r]) = \delta(w)$.*

Proof. By Theorem 3.3, expansion of a distinguished run increases both $L(w)$ and $I(w)$ by one. ■

The *rank* $\rho(w)$ is the number of its distinguished runs. Since there is at most one distinguished run of each type, we have $0 \le \rho(w) \le 3$.

The set of *descendants* $\mathcal{D}(w) \subset \mathcal{G}$ is defined to be the smallest set containing $[w]$ and closed under expansion, in the sense that

$$[v] \in \mathcal{D}(w) \implies [v^+[r]] \in \mathcal{D}(w) \text{ for every distinguished run } r \text{ of } v.$$

For example, if w has rank three, then it has three runs that can be independently extended, and thus its set of descendants has the form

$$\mathcal{D}(w) = \{[w_1(xy)^i w_2(yz)^j w_3(zx)^k] \ : \ i,j,k \ge 0\}.$$

In this case the number of descendants of $[w]$ of length L is the same as the number of ways to express L in the form $L(w) + i + j + k$ with $i,j,k \ge 0$. More generally, keeping in mind our conventions (1.2) and (1.3) on binomial coefficients, we have:

Proposition 5.3 *The number of descendants of w of length L is given by*

$$|\{[v] \in \mathcal{D}(w) \ : \ L(v) = L\}| = \binom{L - L(w) + \rho(w) - 1}{\rho(w) - 1}.$$

Motifs. A word w with $[w] \in \mathcal{G}$ is a *motif* if, for any run r of w with $\ell(r) \geq 4$, there is another run s of the same type with $\ell(r) \leq \ell(s) + 2$. In a motif, there is therefore either a tie or a near–tie for the longest run of a given type, unless there is a unique run r of that type and $\ell(r) \leq 3$.

Let $\mathcal{M} \subset \mathcal{G}$ denote the set of all $[w]$ such that w is a motif.

Proposition 5.4 *Every $[v] \in \mathcal{G}$ is a descendant of a unique motif $[w]$.*

Proof. Given any $[v] \in \mathcal{G}$, repeatedly contract the distinguished runs r of v until either $\ell(r) \leq 3$ or $\ell(r) \leq \ell(s) + 2$ for some run s of the same type as r. The result is a motif $[w] = f([v])$ with $[v] \in \mathcal{D}(w)$. Since we only contract when $\ell(r) \geq 4$, runs other than r are unaffected at each step, and thus the order of contraction does not change the result. It is now readily verified by induction on $L(v)$ that $f([v]) = [w]$ for all motifs $[w]$ and all $[v] \in \mathcal{D}(w)$; hence different motifs have different descendants. ∎

Proof of Theorem 5.1. By Propositions 5.2 and 5.4, the set of $[v] \in \mathcal{G}$ with $\delta(v) = \delta$ is the disjoint union of the descendants of the motifs $[w]$ with $\delta(w) = \delta$; now apply 5.3. ∎

Examples. The 27 motifs $[w] \in \mathcal{G}$ with $\delta(w) = 0$ are listed in Table 8. For brevity we have chosen only one representative motif from each orbit of $\mathrm{Aut}(G)$ acting on $\mathcal{M}$. (Automorphisms of G include, for example, permutations of the generators x, y, z.) The size of the orbit is denoted by $C(w)$, and a representative from the orbit is given in the final column. These 27 motifs correspond to the loops that appear in Figure 3 for $i, j, k \in \{-1, 0, 1\}$.

The 27 motifs with $\delta = 0$ yield formula (1.1) for $N_0(L)$. Similarly, the 12 motifs with $\delta = -1$ yield the formula:

$$N_{-1}(L) = 3\binom{L-2}{1} + 3\binom{L-1}{1} + 6\binom{L-2}{2},$$

and the 153 motifs with $\delta = 1$ yield the formula:

$$N_1(L) = 3\binom{L-5}{-1} + 24\binom{L-5}{0} + 54\binom{L-5}{1} + 12\binom{L-4}{1} + 36\binom{L-5}{2} + 24\binom{L-4}{2}.$$

There are 135 motifs with $\delta = 2$, 603 with $\delta = 3$, 564 with $\delta = 4$, and 2391 with $\delta = 5$.

$\rho(w)$	$L(w)$	$C(w)$	w
0	3	1	$x.y.z.x.y.z$
1	4	6	$x.y.x.y.z.x.y.z$
2	5	6	$x.y.x.y.z.x.y.z.x.z$
2	5	6	$x.y.x.y.z.x.y.z.y.z$
3	6	6	$x.y.x.y.z.x.y.z.y.z.x.z$
3	6	2	$x.y.x.y.z.x.z.x.y.z.y.z$

Table 8. The 27 motifs with $\delta(w) = 0$ come from 6 patterns.

6 Lengths of motifs

In this section we will show that the length of a motif is controlled by its self–intersection number. More precisely, we will prove:

Theorem 6.1 *For any motif w, we have*

$$\delta(w) + \rho(w) + 3 \geq L(w).$$

We then complete our main objective, the proof of Theorem 1.2. We will use the bound above to show that $N_\delta(L)$ is a *finite* sum of binomial coefficients, and that the equality $N_\delta(L) = p_\delta(L)$ holds for all $L \geq \delta + 4$.

Thin motifs. A motif w is *thin* if it has no run with $\ell(r) \geq 4$. The idea of the proof of Theorem 6.1 is to use contractions to reduce to the case of thin motifs. The following quadratic bound plays an important role:

Theorem 6.2 *For any thin motif, we have $I(w) \geq L(w)^2/6$.*

Proof. Under the change of variables $(a = xy, b = yz, c = zx)$, a thin motif becomes a word in (a, b, c) with no repeated letters, so the result follows from Theorem A.1 in the Appendix. ∎

Properties of thin motifs. The conclusion of Theorem 6.1 is equivalent to the bound:

$$\Delta(w) + \rho(w) \geq -3, \tag{6.1}$$

where

$$\Delta(w) = I(w) - 2L(w).$$

The proof is essentially by induction on the length of w. To verify the inequality for motifs of small length, we use the algorithm described in the Appendix to explicitly compute $I(w)$ in many cases. Here are the base cases to be used in the induction.

Theorem 6.3 *The inequality $\Delta(w) + \rho(w) \geq -3$ holds for:*

(a) All thin motifs; and

(b) All motifs with $L(w) \leq 8$.

The sharper inequality $\Delta(w) \geq -3$ holds for:

(c) All thin motifs with $L(w) \geq 6$, and

(d) All imprimitive words that involve all three generators $x, y, z \in \widetilde{G}$.

Proof. Statement (b) is verified by directly calculating $\Delta(w)$ for the 2904 motifs with $L(w) \leq 8$, using the formula for $I(w)$ given in Theorem A.2 below. A similar calculation gives $\Delta(w) \geq -3$ for the 536 thin motifs with $6 \leq L(w) \leq 9$. For $L(w) \geq 10$, Theorem 6.2 implies that any thin motif with $L(w) \geq 10$ satisfies $\Delta(w) \geq 10^2/6 - 20 > -4$; hence $\Delta(w) \geq -3$, and (c) follows. Clearly (c) and (b) imply (a).

To prove (d), note that if w is a *primitive* word involving all three generators x, y, z, then $L(w) \geq 2$, and $I(w) \geq L(w) - 1$ by Theorem 1.1; while for $n \geq 2$, we have $L(w^n) = nL(w)$ and $I(w^n) = n^2 I(w) + n - 1$ by equation (4.1), and therefore

$$\begin{aligned} \Delta(w^n) &\geq (n^2 - 2n)L(w) - n^2 + n - 1 \\ &\geq n^2 - 3n - 1 \geq -3. \end{aligned}$$

■

Remarks. The thin motif $w = (xyxz)^2 yz$ with $L(w) = 5$ satisfies $\Delta(w) = -4$, so statement (c) cannot be sharpened. The imprimitive word $w = (xy)^3$ also satisfies $\Delta(w) = -4$, so the requirement that w involves all three generators is necessary in statement (d).

One can avoid the roughly 3500 special cases implicit in the proof of Theorem 6.3 at the cost of replacing the lower bound of -3 with, say, -10. This would still suffice to prove Theorem 1.2, but with $\delta + 4$ replaced by $\delta + 11$.

Double contraction. To reduce Theorem 6.1 to the case of thin motifs, we will use the following observations about contractions, based on the results of §3.

We will only contract runs with $\ell(r) \geq 4$. This insures that the runs of $w^-[r]$, other than r itself, remain the same as the runs of w. In particular, given two different runs r, s of the same type, we can contract them in either order to obtain the *double contraction* $w^-[r, s]$. By Theorem 3.2, we have:

$$\Delta(w^-[r]) \leq \begin{cases} \Delta(r) + 1 & \text{if } r \text{ is exceptional,} \\ \Delta(r) - 1 & \text{if } r \text{ is not exceptional.} \end{cases}$$

By Theorem 3.1, at most one of the runs r and s is exceptional, so we have:

$$\Delta(w^-[r, s]) \leq \Delta(w).$$

This observation is the crux of the argument.

Proof of Theorem 6.1. Suppose the desired bound is false, and let w be a motif of minimum length with

$$\Delta(w) + \rho(w) < -3. \tag{6.2}$$

Then $L(w) \geq 9$ by Theorem 6.3 part (b), and any contraction w' of w with $\Delta(w') \leq \Delta(w)$ must be primitive by part (d).

Consider the runs of w of a given type, ordered so that $\ell(r_1) \geq \ell(r_2) \geq \cdots \geq \ell(r_n)$. We can assume that the exceptional run, if any, is r_1. By the definition of a motif, either $\ell(r_1) \leq 3$ or $\ell(r_1) \leq \ell(r_2) + 2$.

We claim that $\ell(r_i) \leq 3$ for all $i \geq 3$; otherwise, $w^-[r_i]$ would give a shorter word satisfying equation (6.2). We also have $\ell(r_1) \leq 5$; otherwise, $w^-[r_1, r_2]$ would give a shorter word satisfying (6.2).

We also have $\ell(r_2) \leq 3$, for similar reasons. That is, we cannot have $(\ell(r_1), \ell(r_2)) = (5, 5), (4, 4)$ or $(5, 4)$. Indeed, in the $(5, 5)$ and $(4, 4)$ cases, $w^-[r_2]$ would be a smaller motif satisfying (6.2) (contracting r_2 increases $\rho(w)$ by one, but it also decreases $\Delta(w)$ by one); while in the $(5, 4)$ case, $w^-[r_1, r_2]$ would be a smaller solution.

It follows that, for runs of a fixed type, either $\ell(r_1) \leq 3$ or

$$\ell(r_2) < \ell(r_1) \leq 5.$$

In particular, the number of types with $\ell(r_1) > 3$ is at most $\rho(w)$. By contracting r_1 for each such type, we obtain a thin motif w'. Each contraction reduces the length of w by one, and increases $\Delta(w)$ by at most one; hence

$$\Delta(w) + \rho(w) \geq \Delta(w'),$$

and $L(w') \geq L(w) - \rho(w) \geq 9 - 3 = 6$. Therefore $\Delta(w') \geq -3$ by Theorem 6.3(c), contradicting our assumption (6.2). ■

Remark. The constant 3 in the statement of Theorem 6.1 cannot be improved; for example, the motif $w = xyzxyz$ satisfies $\delta(w) = \rho(w) = 0$ and $L(w) = 3$.

Proof of Theorem 1.2. Fix $\delta \geq -1$. By Theorem 6.1, any motif $[w] \in \mathcal{M}$ with $\delta(w) = \delta$ satisfies

$$L(w) \leq \delta + \rho(w) + 3 \leq \delta + 6,$$

since $\rho(w) \leq 3$. Thus there are only finitely many such motifs, and hence Theorem 1.3 expresses $N_\delta(L)$ as a finite sum of binomial coefficients. Each binomial coefficient has the form

$$b(L) = \binom{L - L(w) + \rho(w) - 1}{\rho(w) - 1}$$

for some $[w] \in \mathcal{M}$. Now recall that $\binom{n}{k}$ agrees with a polynomial of degree k in n for all $n \geq 0$. Theorem 6.1 also implies that

$$L - L(w) + \rho(w) - 1 \geq L - (\delta + 4).$$

Thus for $L \geq \delta + 4$, Theorem 1.3 expresses the value of $N_\delta(L)$ as a quadratic polynomial in L. ■

A Appendix: Counting self–intersections

In this section we give a combinatorial formula for the self–intersection number $I(w)$. This formula was used to produce the data in Table 2. We also obtain the following useful quadratic lower bound on the intersection number:

Theorem A.1 *Let $w \in G$ be a cyclically reduced word in the generators a, b, c and their inverses. Suppose that $L(w) \geq 2$ and no two consecutive letters of w are the same. Then we have $I(w) \geq L(w)^2/6$.*

Group theory. Let $G = \pi_1(M, p)$, where $M = \widehat{\mathbb{C}} - \{0, 1, \infty\}$ and p is a point in the upper halfplane. In this section we will use the presentation

$$G = \langle a, b, c \ : \ abc = \mathrm{id} \rangle \tag{A.1}$$

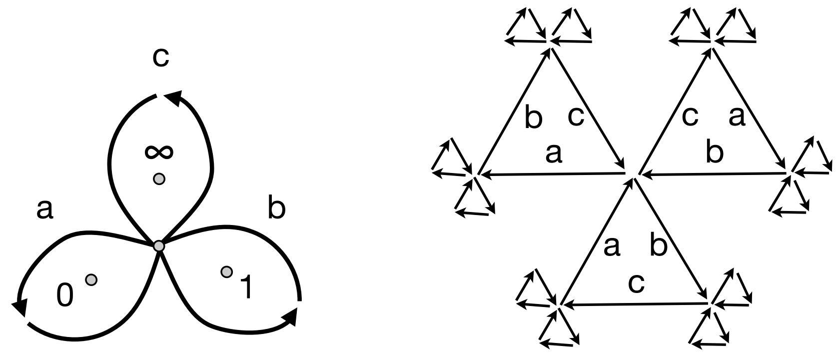

Figure 9. Generators a, b, c for $\pi_1(M)$, and the corresponding Cayley graph H.

for $\pi_1(M,p)$. It is related to the description of G in §2 by the change of variables

$$a = xy, \quad b = yz \quad \text{and} \quad c = zx. \tag{A.2}$$

The presentation (A.1) corresponds to the generating loops for $\pi_1(M,p)$ shown at the left in Figure 9. A portion of the Cayley graph H of G is show at the right. There is a natural realization of H as a graph in the plane, consistent with the ribbon graph structure on H/G inherited from M. The graph H is a proxy for the universal cover $\mathbb{H}$ of M. It carries a canonical metric where each edge has length one.

Intersection numbers. Any two points in H lie on a unique complete geodesic $\delta \subset H$. The intersection number $i(\alpha, \beta)$ of a pair of geodesics in H is defined to be 1 if α and β *cross* in the plane; equivalently, if their endpoints are linked at infinity. Otherwise $i(\alpha, \beta) = 0$. Note that geodesics cannot meet without crossing; in particular, we have $i(\alpha, \alpha) = 0$.

Reduced words. Every $w \in G$ has a unique expression as a *reduced word* in the generators a, b, c and their inverses $\overline{a}, \overline{b}, \overline{c}$. A reduced word has minimal length among all products representing the same element of G. Concretely, this means that consecutive pairs of letters like ab, bc and ca must be avoided, since they can be replaced by $\overline{c}, \overline{a}$ and $\overline{b}$ respectively.

We let $L(w)$ denote the length of w as a reduced word in a, b, c. This notion of length is consistent with the definition $L(w) = \ell(w)/2$ given in §2, as can be seen using equation (A.2).

Stabilizers of geodesics. We say $w \in G$ is *cyclically reduced* if it minimizes $L(w)$ among all elements in its conjugacy class. Concretely, this means that patterns like ab are also avoided when we regard the first and last letters of w as consecutive.

Aside from the identity, every element $w \in G$ stabilizes a unique geodesic $\delta_w \subset H$. This geodesic passes through the identity element $e \in H$ if and only if w is cyclically reduced. If w is primitive, then it generates the stabilizer of δ_w. Similarly, w stabilizes a unique geodesic $\widetilde{\gamma}_w$ in the universal cover $\mathbb{H}$ of M provided $L(w) \geq 2$.

Self–intersection numbers. Now let us fix a primitive, cyclically reduced word

$$w = g_1 g_2 \cdots g_n \in G$$

with $n = L(w) \geq 2$. Its cyclically reduced conjugates are given by

$$w_i = g_i g_{i+1} \cdots g_n g_1 \cdots g_{i-1},$$

$i = 1, 2, \ldots, n$. Let δ_i denote the geodesic through the origin $e \in H$ stabilized by w_i, and let

$$\deg_e(\delta_i \cup \delta_j) = |\{g_i, \overline{g}_{i-1}, g_j, \overline{g}_{j-1}\}| \tag{A.3}$$

denote the number of edges of $\delta_i \cup \delta_j$ incident to the vertex $e \in H$.

Algorithms to compute $I(w) = I(\gamma_w)$ in various situations are well–known; see e.g. [CoL], [DL]. For the case at hand, we will show:

Theorem A.2 *For any primitive, cyclically reduced word $w \in G$, we have*

$$I(w) = \frac{1}{4} \sum_{1 \leq i,j \leq n} i(\delta_i, \delta_j)(\deg_e(\delta_i, \delta_j) - 2). \tag{A.4}$$

Proof. We take as our point of departure the following expression for the self–intersection number as a sum over double cosets:

$$I(w) = \frac{1}{2} \sum_{[g] \in \langle w \rangle \backslash G / \langle w \rangle} i(g \cdot \delta_w, \delta_w). \tag{A.5}$$

This formula is easily justified using hyperbolic geometry, by observing that $g \cdot \widetilde{\gamma}_w$ intersects $\widetilde{\gamma}_w$ in $\mathbb{H}$ iff $i(g \cdot \delta_w, \delta_w) = 1$. Each nonzero term in the sum corresponds to a multiple point $p \in \gamma$ together with an ordered pair of branches of γ through p, and contributes $1/2$ to the total intersection number $I(\gamma_w) = I(w)$.

To connect formulas (A.5) and (A.4), suppose that $i(g \cdot \delta_w, \delta_w) = 1$. Then

$$(g \cdot \delta_w) \cap \delta_w$$

is a compact geodesic interval in H with endpoints $h_1, h_2 \in G$. Translating by h_1^{-1}, we obtain a pair of geodesics passing through e, each stabilized by a conjugate of w; hence

$$(h_1^{-1} g \cdot \delta_w, h_1^{-1} \cdot \delta_w) = (\delta_i, \delta_j)$$

for some $1 \leq i, j \leq n$. The intersection number is preserved by translation, so $i(\delta_i, \delta_j) = 1$.

If $h_1 = h_2$ then δ_i and δ_j cross transversally at e; hence $\deg_e(\delta_i \cap \delta_j) = 4$ and we obtain a contribution of $1/2$ to the sum in formula (A.4). If $h_1 \neq h_2$ then the degree is 3, and we obtain two terms, one from h_1 and one from h_2, each contributing $1/4$ to the same sum. All the nonzero terms in equation (A.4) arise in this way, and hence the two sums are equal. ■

Algorithmic considerations. The term $i(\delta_i, \delta_j)$ appearing in formula (A.4) can be readily computed by comparing the cyclic orderings of the edges of δ_i and δ_j at the endpoints of the segment $\delta_i \cap \delta_j$. The degree is computed by equation (A.3).

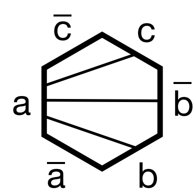

Figure 10. Chords represent geodesics passing through the origin $e \in H$.

Proof of Theorem A.1. Let $\mathcal{A} = \{a, b, c, \overline{a}, \overline{b}, \overline{c}\} \subset G$ be the alphabet consisting of the generators of G and their inverses. We can also regard $\mathcal{A}$ as the cyclically ordered set of vertices of H adjacent to e. Let P be a hexagon whose sides are labeled by the elements of $\mathcal{A}$ in the same cyclic order; see Figure 10. A *chord* $[u, v]$ is an unordered pair of distinct elements $u, v \in \mathcal{A}$ labeling non-adjacent sides of P. For example, the chords incident to side a are given by $[a, b]$, $[a, \overline{b}]$ and $[a, c]$.

The set of all chords K consists of 9 elements. We will also regard the set K as a basis for the vector space $\mathbb{R}^K$. Define a symmetric bilinear form on $\mathbb{R}^K$ by $Q(k, k') = 1$ if the chords k and k' cross in P, and $Q(k, k') = 0$ otherwise. (By definition, two chords incident to the same side of P do not cross.) It is easily checked that Q is an indefinite form of signature $(6, 3)$, by diagonalizing the given matrix.

Let $S = \{a, b, c\}$, and let

$$\partial : \mathbb{R}^K \to \mathbb{R}^S$$

be the linear boundary map that counts $+1$ when a generator occurs as an endpoint of a chord, and -1 when its inverse occurs. (For example, $\partial([a,b]) = a + b$, while $\partial([a,\overline{b}]) = a - b$.) Let

$$\lambda : \mathbb{R}^K \to \mathbb{R}$$

be the length function defined by summing the coordinates; that is, by setting $\lambda(k) = 1$ for each chord k. One can now readily verify that $Q|\operatorname{Ker}(\partial)$ is positive–definite, and that for any vector $v \in \operatorname{Ker}(\partial)$, we have

$$Q(v, v) \geq \lambda(v)^2/3. \tag{A.6}$$

(The minimum is achieved on vectors that give weight zero to all chords that join opposite sides, and equal weights to the rest.)

Now let $w = g_1 \cdots g_n \in G$ be a primitive, cyclically reduced word such that no pair of adjacent letters (g_{i-1}, g_i) are equal. (Here g_n is adjacent to g_1.) This condition, plus the fact that w is reduced, insures that the sides of P labeled by $\overline{g}_{i-1}$ and g_i are not adjacent. Thus we may associate to w the sequence of chords

$$k_i = [\overline{g}_{i-1}, g_i],$$

$i = 1, \dots, n$.

Let $v = \sum_1^n k_i \in \mathbb{R}^K$. Then clearly $\overline{\partial}(v) = 0$ and $\lambda(v) = L(w) = n$. Using formula (A.4), it is also easy to see that

$$I(w) \geq \frac{1}{2} \sum_{1 \leq i,j \leq n} Q(k_i, k_j) = \frac{1}{2} Q(v, v).$$

Indeed, when $Q(k_i, k_j) = 1$ the geodesics δ_i and δ_j cross transversally at the identity element $e \in H$, with with degree 4 in the sense of (A.3), and hence we obtain a contribution of $1/2$ to the sum in formula (A.4). Referring to inequality (A.6), we obtain the desired inequality

$$I(w) \geq \lambda(v)^2/6 = L(w)^2/6.$$

■

References

[Bon] F. Bonahon. The geometry of Teichmüller space via geodesic currents. *Invent. math.* **92**(1988), 139–162.

[CL] M. Chas and S. P. Lalley. Self-intersections in combinatorial topology: statistical structure. *Inv. math.* **188**(2012), 429–463.

[CP1] M. Chas and A. Phillips. Self-intersection numbers of curves on the punctured torus. *Exp. Math.* **19**(2010), 129–148.

[CP2] M. Chas and A. Phillips. Self-intersection numbers of curves in the doubly punctured plane. *Exp. Math.* **21**(2012), 26–37.

[CoL] M. Cohen and M. Lustig. Paths of geodesics and geometric intersection numbers I. In *Combinatorial Group Theory and Topology, Alta, Utah, 1984*, volume 111 of *Annals of Math. Studies*, pages 479–500. Princeton Univ. Press, 1987.

[dGS] M. de Graaf and A. Schrijver. Making curves minimally crossing by Reidemeister moves. *J. Combin. Theory Ser. B* **70**(1997), 134–156.

[DL] V. Despré and F. Lazarus. Computing the geometric intersection number of curves. *Preprint, 2016.*

[Er] V. Erlandsson. A remark on the word length in surface groups. *Preprint, 2016.*

[EPS] V. Erlandsson, H. Parlier, and J. Souto. Counting curves, and the stable length of currents. *Preprint, 2016.*

[ES] V. Erlandsson and J. Souto. Counting curves in hyperbolic surfaces. *Geom. Funct. Anal.* **26**(2016), 729–777.

[FLP] A. Fathi, F. Laudenbach, and V. Poénaru. *Travaux de Thurston sur les surfaces.* Astérisque, vol. 66–67, 1979.

[HS] J. Hass and P. Scott. Intersections of curves on surfaces. *Israel J. Math.* **51**(1985), 90–120.

[IT] Y. Imayoshi and M. Taniguchi. *An Introduction to Teichmüller Spaces.* Springer-Verlag, 1992.

[Mir1] M. Mirzakhani. Growth of the number of simple closed geodesics on hyperbolic surfaces. *Annals of Math.* **168**(2008), 97–125.

[Mir2] M. Mirzakhani. Counting mapping class group orbits on hyperbolic surfaces. *Preprint, 2016.*

[Re] B. L. Reinhart. Algorithms for Jordan curves on compact surfaces. *Annals of Math.* **75**(1962), 209–222.

[St] J. Stillwell. *Geometry of Surfaces.* Springer-Verlag, 1992.