The geometry and combinatorics of closed geodesics on hyperbolic surfaces

by

Chris Arettines

Abstract

The geometry and combinatorics of closed geodesics on hyperbolic surfaces

by

Chris Arettines

Advisor: Professor Ara Basmajian

In this thesis, we obtain combinatorial algorithms that determine the minimal number of self-intersections necessary for a free homotopy class $[\gamma]$ on an orientable surface, using algebraic input. Using this same input, we describe another algorithm which determines whether or not a minimally intersecting curve in $[\gamma]$ is *filling*, that is, whether or not the complement is a disjoint union of disks or punctured disks. Next, we use these algorithms as inspiration for proving the existence of filling curves which self-intersect $2g-1$ times, which is the minimal number of intersections possible. The combinatorial viewpoint that is developed can then be used to obtain geometric information about the curves, which is the subject of the last chapter. Among other things, we obtain a sharp lower bound on the length of a filling curve with the minimal number of self-intersections on a surface of genus g.

Contents

List of Figures

Chapter 1

Overview

In this thesis we analyze the interplay between algebra, combinatorics, geometry, and topology in the context of studying closed geodesics on orientable surfaces which carry a hyperbolic metric. A surface can be encoded as an edge-gluing pattern X for a fundamental polygon, and a free homotopy class can be specified by an *edge-crossing sequence* W in that fundamental polygon(see Chapter 2). Once a hyperbolic metric is chosen for the surface, every free homotopy class contains a unique geodesic. On surfaces which carry a metric of negative curvature (which are all that will be considered here), there are certain topological invariants of closed geodesics (or free homotopy classes) which are independent of the particular metric chosen. The first invariant is the *geometric intersection number*, which is the minimal possible number of crossings that the free homotopy class has on the surface. A closed geodesic always has the minimal number of intersections. The next invariant is whether or not the geodesic is *filling*, that is, whether or not the complement of the geodesic is a disjoint union of disks or once punctured disks. Since the intersection number and filling property are invariants

of the geodesic, one should be able to determine these properties using the algebraic information of X and W. There is a long history of studying these properties from this viewpoint (see [10],[16],[12] for example). In Chapter 3 we will present some new methods to determine the intersection and filling properties of a free homotopy class.

Using a basic Euler characteristic argument, one can prove that any filling curve on a surface of genus g must have at least $2g-1$ intersections. One can then ask if such curves actually exist. In [4], it is proven that there are no *pairs* of simple curves (curves without self-intersections) which fill on a surface of genus 2, while there are pairs for higher genus surfaces. In Chapter 4, we show that for every surface of genus ≥ 2, there are self-intersecting filling curves which have $2g-1$ self-intersections. We also show that there are filling curves with $2g$ self-intersections, and conjecture that there is a filling curve with k intersections, for every $k \geq 2g-1$. Because the complements of filling curves are topologically simple, they are amenable to a combinatorial analysis, and the construction in Chapter 4 hints at a richer structure awaiting further study.

In Chapter 5, we present a very general point of view for analyzing filling geodesics by means of their combinatorics. Suppose we fix a hyperbolic metric on a surface S. Then each filling curve γ decomposes S into a collection of hyperbolic polygons, whose angles and side lengths are organized according to the combinatorics of γ. For certain γ which lack so-called *triangular regions*, these combinatorics are independent of the particular hyperbolic metric chosen for S. Facts about hyperbolic polygons can then be used to obtain information about the length of γ. Using this viewpoint, we obtain a sharp lower bound on the length of any minimally intersecting filling curve on a surface of genus g. In the part of Chapter 5, we analyze the role that

angles of intersection play as a hyperbolic metric is varied on the punctured torus. In [30] it is shown that there are three closed geodesics whose pairwise angles of intersection parameterize the hyperbolic structure. We exhibit a pair of curves whose angles of intersection *fail* to parameterize the metric, and attempt to precisely quantify the failure. This result represents a first step in a wider study of understanding how properties of geodesics (angles of intersection, lengths) interact with the overall hyperbolic metric on a surface.

Chapter 2

Basic Geometry and Topology

In this chapter, we will take a whirlwind tour of the basics of surface topology and geometry, with an emphasis on the concepts that will be used implicitly and explicitly through the rest of this thesis. The reader may consult [2], [6],[11],[23],[28],[29], or [35] for excellent expositions of these subjects.

2.1 Hyperbolic Geometry

There are three standard geometries in two dimensions, namely Euclidean, spherical and hyperbolic geometries. Hyperbolic geometry turns out to the most prolific of these geometries in two dimensions, and thus we begin by recalling some of the basic facts of hyperbolic geometry that underpin the discussion that follows in later chapters.

2.1.1 Models of the Hyperbolic Plane

There are various representations of the hyperbolic plane that sit inside of Euclidean space. One such model is the so-called *Upper Half-Plane Model*. Let $\mathbb{U} = \{z \in \mathbb{C} | Im(z) > 0\}$ and endow $\mathbb{U}$ with the metric induced by the infinitesimal arc-length element $\frac{|dz|}{Im(z)}$, where $|dz|$ is the normal Euclidean arc-length element. $\mathbb{U}$ equipped with this arc-length element is called the *upper half-plane model of the hyperbolic plane*. The length of any piecewise differentiable path $\gamma : [0, 1] \to \mathbb{U}$ is obtained by integrating over the path in the usual way:

$$l(\gamma) = \int_0^1 \frac{|\gamma'(t)|}{Im(\gamma(t))} dt$$

The distance between two points z_1 and z_2 is then given as the infimum over all $\mathcal{C}^1$ paths in $\mathbb{U}$ connecting z_1 to z_2. The set $\mathbb{U}$ with this induced distance $d_{\mathbb{U}}$ is the *upper half-plane model* of hyperbolic geometry. The set $\partial\mathbb{U} = \{z \in \mathbb{C} | Im(z) = 0\} \cup \{\infty\}$ is called the *boundary at infinity* of the *boundary* of the hyperbolic plane.

From the formula for arc-length in $\mathbb{U}$, it can be shown that geodesics in $\mathbb{U}$ are either vertical Euclidean half-lines or semi-circles orthogonal to $\partial\mathbb{U}$. There are many other important properties of geodesics that follow from this fundamental observation about geodesics, which we summarize:

Proposition 2.1.1. ***Facts about geodesics***

1. *There is a unique geodesic passing through any two distinct points in* $\mathbb{U}$.

2. *A geodesic is uniquely determined by its endpoints on* $\partial\mathbb{U}$.

3. *If L_1 and L_2 are geodesics with distinct endpoints on $\partial\mathbb{U}$, then there is a unique geodesic O which is orthogonal to L_1 and L_2.*

Any homeomorphism between $\mathbb{U}$ and another set $X \subset \mathbb{C}$ induces a metric on X via pullback. The map $g(z) = \frac{z-i}{z+i}$ is a bijective conformal map which sends $\mathbb{U}$ to $\mathbb{D} = \{z \in \mathbb{C}—|z| < 1\}$. Thus, $\mathbb{D}$ inherits a metric $d_{\mathbb{D}}$, which has the arc-length element $\frac{2|dz|}{1-|z|^2}$. The unit disk $\mathbb{D}$ with this induced metric $d_{\mathbb{D}}$ is called the *unit disk model of the hyperbolic plane*. The boundary at infinity $\partial\mathbb{D}$ in this case is just the unit circle S^1. Since $g(z)$ is conformal, one sees that geodesics in this model are Euclidean diameters of S^1, along with arcs of Euclidean circles in $\mathbb{C}$ which intersect S^1 orthogonally. The properties given in Proposition 2.1.1 also carry over.

Even though the two models discussed are isometric, they provide a slightly different perspective for thinking about hyperbolic geometry. Certain insights become more apparent when using the right model. There are other models for the hyperbolic plane which we will not discuss here, but the reader may refer to the earlier references or [37] for additional details and a unifying perspective.

In the arguments that follow, if the particular model is not important, or is implicitly understood, we will simply use the notation $\mathbb{H}^2$ and $\partial\mathbb{H}^2$ to refer to the hyperbolic plane and its boundary.

2.1.2 Isometries

An *orientation-preserving isometry* of a metric space X is a conformal bijective map f from X to itself such that $d(x, y) = d(f(x), f(y))\ \forall x, y \in X$. Since $\mathbb{U}$ and $\mathbb{D}$ can be realized as subsets of the Riemann sphere $\hat{\mathbb{C}}$, every orientation-preserving isometry of either model can be extended to a con-

formal bijective map on all of $\hat{\mathbb{C}}$ using the maps $z \mapsto \bar{z}$ or $z \mapsto \frac{1}{\bar{z}}$. Thus, the group of orientation preserving isometries of either model must be isomorphic to subgroups of the *Möbius group*, the group of conformal bijections of the Riemann sphere. The following theorem is classical and points the way to a full classification of isometries:

Theorem 2.1.2. *Every conformal automorphism of* $\hat{\mathbb{C}}$ *can be written in the form:*

$$f(z) = \frac{az+b}{cz+d}, \text{ where } ad - bc = 1.$$

From this, we can deduce the following:

Theorem 2.1.3. ***Descriptions of Orientation-preserving Isometries***

1. *Every orientation preserving isometry* $f(z)$ *of* $\mathbb{U}$ *can be written in the form:*

$$f(z) = \frac{az+b}{cz+d}, where\ a, b, c, d \in \mathbb{R}\ and\ ad - bc = 1.$$

2. *Every orientation preserving isometry* $f(z)$ *of* $\mathbb{D}$ *can be written in the form:*

$$f(z) = \frac{az+\bar{c}}{cz+\bar{a}}, where\ a, c \in \mathbb{C}\ and\ |a|^2 - |c|^2 = 1$$

Proof. f is conformal and bijective, and sends $\mathbb{U}$ to itself since all of the coefficients are real. A simple calculation shows that $\frac{|f'(z)|}{Im(f(z))} = \frac{|dz|}{Im(z)}$. Conjugating by the conformal map $g(z)$ from the previous section yields the 2nd description. □

There is a natural map ϕ from the group of orientation preserving isometries of $\mathbb{H}^2$ to $GL(2, \mathbb{R}$ given by realizing the isometry as a rational map

$f(z) = \frac{az+b}{cz+d}$ acting on $\mathbb{U}$ via Theorem 2.1.3, and sending it to the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. It is an incredible fact that the composition of two rational maps $f(z) = \dfrac{az+b}{cz+d}$ and $g(z) = \dfrac{a'z+b'}{c'z+d'}$ behaves like multiplication of the matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right)$, so ϕ is actually a homomorphism. This yields the following algebraic description of the orientation-preserving isometry group of the hyperbolic plane.

Theorem 2.1.4. *$Isom^+(\mathbb{H}^2) \simeq PSL(2,\mathbb{R})$, which is the group of 2×2 matrices with determinant 1 modulo the subgroup $K = \left(\begin{smallmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{smallmatrix}\right)$*

Proof. ϕ maps to $PGL(2,\mathbb{R})$ by Theorem 2.1.3. Multiplying all of the coefficients of a rational map by the same number does not change the rational map, but does change the image matrix. Since we require the image matrix to have determinant 1, we can multiply each coefficient by -1 and obtain a different matrix corresponding to the same rational map. Thus the kernel of ϕ is precisely $K = \left(\begin{smallmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{smallmatrix}\right)$. □

The isometries of the hyperbolic plane fall into three categories, based on the number and location of the fixed points of the isometry.

Theorem 2.1.5. ***Classification of isometries*** *Let f be a non-trivial orientation preserving isometry of the upper half-plane model of $\mathbb{H}^2$. Then exactly one of the following holds:*

1. *f has exactly one fixed point in $\mathbb{U}$ and is conjugate to the map $z \mapsto e^{i\theta}z$ for some $\theta \in \mathbb{R}$, in which case f is said to be **elliptic**.*

2. *f has exactly one fixed point in $\partial\mathbb{U}$ and is conjugate to the map $z \mapsto z+1$, in which case f is said to be **parabolic**.*

3. *f has exactly two fixed points in $\mathbb{U}$ and is conjugate to the map $z \mapsto \lambda z$ for some $\lambda > 0, \lambda \neq 1$, in which case f is said to be **hyperbolic**.*

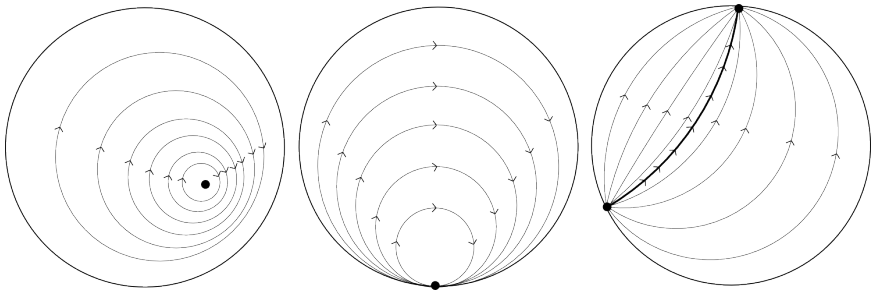


Figure 2.1: Elliptic, parabolic, and hyperbolic isometries respectively.

2.1.3 Trigonometry

There are hyperbolic trigonometric identities analogous to their Euclidean counterparts. In addition, there is a second law of cosines for hyperbolic triangles which implies that there are no similarities in hyperbolic geometry. A proof of the formulas below can be found in [6].

Proposition 2.1.6. ***Hyperbolic Trigonometric Identities*** *Let Δ be a hyperbolic triangle with angles α,β,γ and corresponding side lengths A,B,C. Then the following relationships hold:*

1. $\frac{sinh(A)}{sin(\alpha)} = \frac{sinh(B)}{sin(\beta)} = \frac{sinh(C)}{sin(\gamma)}$
2. $cosh(C) = cosh(A)cosh(B) - sinh(A)sinh(B)cos(\gamma)$
3. $cosh(C) = \frac{cos(\alpha)cos(\beta) + cos(\gamma)}{sin(\alpha)sin(\beta)}$

2.2 Curves and Surfaces

The spaces of interest in this thesis will for the most part be two-dimensional orientable surfaces, orientable manifolds which are locally homeomorphic to $\mathbb{R}^2$. We will also consider surfaces with punctures, which are surfaces with

a finite number of points removed. Details of the viewpoint established here can be found in [34] or [35].

Theorem 2.2.1. *Orientable surfaces (with or without punctures) are classified up to homeomorphism by their genus and number of points removed.*

We will sometimes use the notation $S_{g,n}$ to denote a topological surface of genus g with n punctures. When $n = 0$, we will simply write S_g. Recall that the *fundamental group* of a connected topological space X, denoted $\pi_1(X)$, is the group of homotopy classes of closed curves based at a chosen point $x \in X$. The fundamental groups of surfaces have many interesting properties that could fill many volumes. For now, we will content ourselves with an algebraic presentation of the particular surfaces that will be of concern in this thesis:

Theorem 2.2.2. *Let $S_{g,n}$ be a topological surface of genus g with n punctures satisfying $n + 2g \geq 3$ Then:*

1. *If $n > 0$, $\pi_1(S_{g,n}) \simeq \mathbb{F}_{2g+n-1}$, where $\mathbb{F}_m$ is the free group on m generators.*

2. *If $n = 0$, $\pi_1(S_{g,0}) \simeq \langle a_1, b_1, ..., a_g, b_g | \prod_{i=1}^{g} [a_i, b_i] = 1 \rangle$*

The fundamental group of a space captures the algebraic characteristics of closed loops that are *based* at a particular point. By a If we loosen this restriction to look at the set of *free homotopy classes*, we obtain the following basic topological fact:

Proposition 2.2.3. *Free homotopy classes of closed curves in a space X are 1-1 correspondence with conjugacy classes of elements in $\pi_1(X)$*

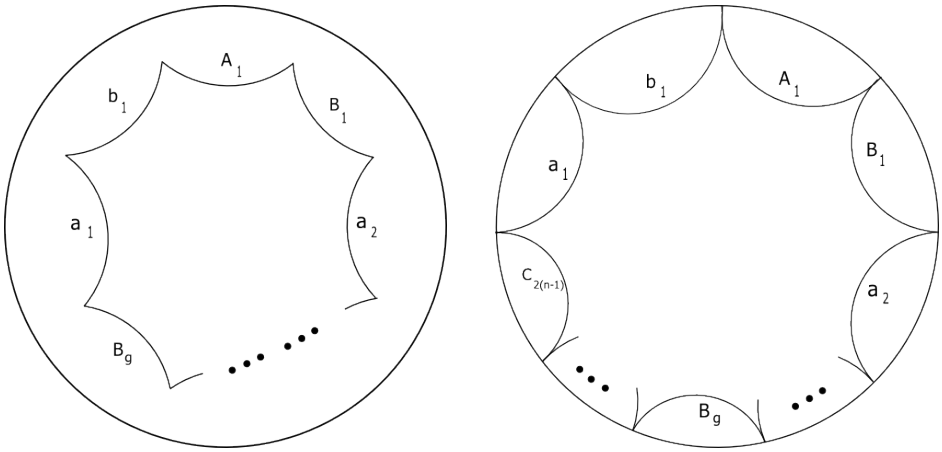
Free homotopy classes are much more natural to consider in geometric settings than based homotopy classes.

Remark 2.2.4. On Notation:

- We use often use the word *curve* to mean a closed curve. We will call a collection of curves a *multicurve*. Curves and multicurves may possibly have many intersections as they weave around the surface.
- From now on, we will use capital letters to indicate inverses of generators of π_1.

2.2.1 Surfaces and hyperbolic geometry

A topological surface of genus g can be obtained by taking a $4g$-sided polygon and identifying the edges as shown in Figure 2.2a so that all of the corners meet up to form a single identified vertex. If we use Euclidean polygons, then if $g \geq 2$ the total angle at the vertex will exceed 2π, so we cannot induce a smooth Euclidean structure on the surface in this way. However, in hyperbolic geometry, there exist regular $4g$-sided polygons whose angles are all $\pi/2g$, so that the corners of the polygon smoothly meet up at the vertex. One can realize the gluing identifications using hyperbolic isometries, which then induces a hyperbolic metric on the underlying surface. A surfaces with punctures can be realized by considering *ideal polygons* with vertices on $\partial\mathbb{H}^2$. If $n \geq 1$, then $S_{g,n}$ (with $n+2g \geq 3$) can be obtained from an ideal hyperbolic polygon with $4g + 2(n-1)$ sides as in Figure 2.2b. The metrics constructed in this way are *complete* and yield surfaces with *finite area*.



(a) A schematic fundamental polygon for S_g

(b) A schematic fundamental polygon for $S_{g,n}$

Figure 2.2: Geometric realizations of fundamental polygons

A polygon constructed as above will be called a *standard fundamental polygon* for a surface $S_{g,n}$. Since the edge identifications are realized using hyperbolic isometries, we can tessellate all of $\mathbb{H}^2$ with isometric copies of the initial polygon. In other words, $\mathbb{H}^2$ is the *universal cover* of our surface. The group of gluing isometries is the *deck group* of this covering, and is thus isomorphic to the fundamental group of our surface. A detailed account of this body of ideas can be found in [28].

Given a standard fundamental polygon whose edges are labeled $a_1, b_1, ..., a_g, b_g, A_g, B_g, c_1, C_1, ..., c_{2(n-1)}, C_{2(n-1)}$ as in Figure 2.2, there is a natural way to visualize a set of generators for $\pi_1(S_{g,n})$. Let the center of the polygon be a basepoint, and let the generating closed loops be the paths that start at the basepoint, cross through a single edge of the polygon, and return to the basepoint. Each such loop can be identified with the edge that it crosses. Using this set of generators (which is not the set of generators

typically used in introductory topology books such as [23] or [29]), we obtain a slightly different presentation for the fundamental groups of the surfaces S_g. The set of generators is the same, but the defining relation becomes $\prod_{i=1}^{g}[B_i a_i] = 1$.

A *closed geodesic* is a curve which has minimal length within its free homotopy class. On a Euclidean torus, each free homotopy class contains an infinite number of closed geodesic curves. The opposite is true for surfaces which carry a hyperbolic metric:

Theorem 2.2.5. *For surfaces with a hyperbolic metric, every free homotopy class of closed curves contains a unique geodesic*

A detailed proof of this can be found in [17], but it is essentially due to the fact that the lifts of a curve converge to two distinct points along $\partial\mathbb{H}^2$. These two points determine a geodesic in $\mathbb{H}^2$ which is a lift of the geodesic representing the same free homotopy class.

If a surface S has punctures, then each free homotopy class can be associated to an *edge crossing word*, which is unique up to conjugation. If the surface does not have punctures, then the edge crossing word is unique up conjugation and swapping out complementary halves of the defining relation.

2.3 The Mapping Class Group and Teichmüller Space

Two homeomorphisms ϕ_0 and ϕ_1 of a space S are said to be *isotopic* if there is a map $H : [0,1] \times S \to S$ such that $H(0,x) = \phi_0(x)$ and $H(1,x) = \phi_1(x)$ $\forall x \in S$, with $H(t,*)$ a homeomorphism for each $t \in [0,1]$. The group of homeomorphisms modulo this isotopy equivalence is called the *mapping class*

group of S. The mapping class groups of surfaces have deep connections to the theory of 3-manifolds and are also interesting in their own right. An excellent introduction to the subject can be found in [1],[18] or [17]. We will eventually need the following fact related to mapping class groups:

Proposition 2.3.1. *Let $\phi : \mathbb{D} \to \mathbb{D}$ be a homeomorphism of the unit disk which fixes the boundary S^1 pointwise and fixes the origin. Then ϕ is isotopic to the identity homeomorphism.*

This will be used in Chapter 4 to analyze curves on surfaces.

We have seen in the previous section that many surfaces S can be given a hyperbolic metric. We can obtain a new metric on S by applying a homeomorphism and using the pullback metric. In certain settings, we would like to distinguish these two metrics on S, while in other situations, we may want to consider them equivalent. Following this line of thought leads to the following definition:

Definition 2.3.2. The *Teichmúller space* of a surface $S = S_{g,n}$, denoted $Teich(S)$, is the space of equivalence classes $\{(X, \phi)\}/\sim$, where X is a surface with a complete finite area hyperbolic metric, $\phi : S \to X$ is a homeomorphism, and $(X', \phi') \sim (X, \phi)$ if $\phi' \circ \phi^{-1}$ is homotopic to an isometry.

Teichmüller space captures the different ways in a which an abstract topological surface S can "wear" a hyperbolic metric. The map ϕ is called the *marking* of the surface. The mapping class group of S naturally acts on Teichmüller space by precomposing ϕ by a homemorphism within an isotopy class. Teichmüller spaces are homeomorphic to open balls, and carry many interesting metrics which are still not fully understood.

2.3.1 Fenchel-Nielsen coordinates

Here we will briefly introduce the so called *Fenchel-Nielsen* coordinates on Teichmüller space, which provide a way of making the abstract definition of Teichmüller more concrete. We begin by cutting a hyperbolic surface into so called *pairs of pants*, which are topologically spheres with a combination of three holes or punctures. This can be done by choosing a maximal collection of disjoint non-isotopic closed geodesics with no self-intersections. Closed curves with no self-intersections will be called *simple curves*.

Proposition 2.3.3. *The maximal number of disjoint non-isotopic simple curves on a surface $S_{g,n}$ is $3g - 3 + n$.*

Proof. First, assume that $n = 0$. First, notice that every simple closed curve on a pair of pants is isotopic to one of the boundary components. So if a collection of curves cuts a surface into pairs of pants, the collection of curves must be maximal. Since the Euler characteristic of a circle is 0 and the Euler characteristic of a pair of pants is -1, we would need $2g-2$ pairs of pants to reassemble a surface S_g. The $6g - 6$ closed curves we obtain from these disjoint pairs of pants are identified in pairs, giving us $3g - 3$ closed curves on our original surface as desired. Adding a puncture turns a pair of pants into a four-holed sphere, which can be cut in half by a simple curve to obtain two pairs of pants. □

Proposition 2.3.4. *A hyperbolic metric on a topological pair of pants is completely determined by the lengths of the three boundary components (let punctures have length 0).*

This proposition can be proven by using a slightly modified version of the following lemma:

Lemma 2.3.5. *Let $a, b, c > 0$ be given. Then there is a unique right-angled hyperbolic hexagon with alternating side lengths a, b, c.*

A hyperbolic pair of pants can be constructed from such a pair of right-angled hexagons. If one of more of the lengths in Proposition 2.3.4 are 0, then we use an analogous lemma for hyperbolic pentagons, quadrilaterals, or triangles with 1, 2, or 3 ideal vertices respectively to yield the desired geometric object. An *orthosegment* is a portion of a hyperbolic geodesic which meets one or more other geodesics orthogonally. From the previous arguments, we see the following:

Corollary 2.3.6. *There is a unique orthosegment connecting any two boundary components of a hyperbolic pair of pants.*

Now suppose that we are given a point in Teichmüller space, that is, a pair (X, ϕ). Choose a collection of $3g-3+n$ oriented curves $P = \{\gamma\}_i$ which cut $S_{g,n}$ into topological pairs of pants.

Also choose a collection of *seams* $Q = \{\delta\}_j$ for P, that is, a set of disjoint oriented simple curves such that the intersection of any $\delta \in Q$ with any pair of pants determined by P is a union of three arcs connecting the boundary components pairwise.

The curves in P get mapped to X via ϕ, and thus each curve has a unique geodesic representative. These $3g - 3 + n$ lengths are called the *length parameters* for the point in Teich(S). Take this collection of closed geodesics, and consider their full set of lifts to the universal cover $\mathbb{H}^2$. Also consider the full set of lifts of the geodesic representatives of the curves in Q. We will now define the twist parameter about a curve γ_i in P. Consider a fixed lift of γ_i, $\tilde{\gamma_i}$, and consider a fixed lift of a seam curve δ_j, $\tilde{\delta_j}$,which intersects $\tilde{\gamma_i}$. Since $\tilde{\delta_j}$ is a lift of a seam curve, it intersects an infinite number

of lifts of the curves in P on either side of $\tilde{\gamma_i}$. Let X and Y be the adjacent geodesic lifts that $\tilde{\delta_j}$ crosses on each side of $\tilde{\gamma_i}$ on the right and left sides of $\tilde{\gamma_i}$ respectively. This is well defined since γ_i is oriented. Let O_X and O_Y be the orthogonal segments between X and $\tilde{\gamma_i}$, Y and $\tilde{\gamma_i}$ respectively. The *twist parameter about* γ_i is the oriented distance between the basepoints of O_X and O_Y along $\tilde{\gamma_i}$. One has to check that this definition is meaningful (i.e., doesn't depend on the choices of lifts) which is routine. It can also be shown that this process can be reversed: namely, given a pants composition and collection of seams P and Q along with $3g-3+n$ length parameters and $3g-3+n$ twist parameters, a point in $Teich(S_{g,n})$ can be uniquely defined. We thus arrive at the following theorem:

Theorem 2.3.7. *$Teich(S_{g,n})$ is homeomorphic to $\mathbb{R}_+^{3g-3+n} \times \mathbb{R}^{3g-3+n}$.*

Corollary 2.3.8. *$Teich(S_{1,1})$ is homeomorphic to the unit disk $\mathbb{D}$.*

Even more is true: Every Teichmüller space inherits a natural metric called the *Teichmüller metric*, which we will not discuss here (see [1],[18],[17] for good expositions), and in the case of the punctured torus, this metric agrees with the hyperbolic metric.

Chapter 3

Combinatorial Algorithms

The 1-1 correspondence between geodesics and conjugacy classes of words in the fundamental group suggests that certain geometric information about a closed geodesic can be extracted from a word representing a conjugacy class. In this section, we will survey some of the literature exploring this concept, and present some algorithms which solve topological and geometric problems using algebraic input.

3.1 Minimal Configurations - Two algorithms

One of the most basic topological concepts that can be attached to a curve or collection of curves, is the notion of intersection.

Definition 3.1.1. The *geometric intersection number* of a primitive free homotopy class is the minimal possible number of transverse intersections of a curve in the class.

The geometric intersection number for a *pair* of curves or any other larger group of curves is defined similarly. In this section, we will consider

two separate methods which actually produce a visual representative of a free homotopy class which has the minimal possible number of intersections. There are two slightly different problems we consider:

1. If we are given a collection of arcs in a fundamental polygon, how do we determine if it is a configuration with the minimal number of intersections?

2. Given a free homotopy class described by an edge-crossing sequence for some fundamental polygon, can we construct a minimally intersecting representative in an efficient way?

Algorithms for tackling the intersection problem can be found in [10],[12],[13],[15], and [16],[21],[33] using various input data. The algorithms presented in [10] and [16] use the same input data described here, and provide the motivating framework for the current discussion.

We will describe two approaches to the intersection problems described above. The first approach is "greedy" in the sense that it first constructs an arbitrary representative, and then analyzes the combinatorial structure to determine if there are any excess intersections. It then modifies the representative appropriately. This method provides an answer to the first question above.

The second approach constructs a representative which is guaranteed to have the minimal number of intersections at the outset. We include both methods to demonstrate the rich combinatorial structure that these basic topological objects can carry. The first method was originally worked out under the guidance of Moira Chas. The presentation here is a newer, cleaner version of the original arguments, while the second approach is new.

The crucial fact is that if a representative has excess self-intersections, then the curve is guaranteed to have a collection of *bigons* or *monogons*. This is a theorem of Hass and Scott that seems intuitively obvious, but is in fact difficult to prove.

Definition 3.1.2. A smooth map $f : S^1 \to S_g$ has a *monogon* if there is an immersed disk in S_g whose boundary consists of the image of a subarc $[a, b]$ of S^1 with $f(a) = f(b)$. f is has a *bigon* if there is an immersed disk in S_g whose boundary consists of the image of two subarcs $[a, b]$ and $[c, d]$ of S^1 such that $f(a) = f(c)$ and $f(b) = f(d)$. The bigon is said to be *proper* if the two subarcs are disjoint.

Theorem 3.1.3. *([20]) If a representative has excess self-intersection, then it contains a proper bigon or monogon.*

The first algorithm we describe will based on a combinatorial analogue of Theorem 4.1.3. To motivate the algorithm, we will first look at an example.

3.1.1 A Motivating Example

Consider the surface $S_{2,1}$, a surface of genus 2 with 1 puncture. It can be represented topologically as an octagon with the vertices of the polygon removed. Consider the free homotopy class corresponding to the word $a^2cdB^3 \in \mathbb{F}_4$. We draw the octagon with the edge labels, and then construct any curve in $[\gamma]$ by tracing out a curve with edge-crossing sequence a^2cdB^3. We want to see if this curve we've drawn has the minimal number of intersections possible among all curves in $[\gamma]$. One way to modify our curve to remove obvious excess intersections is to straighten out the arcs crossing the interior of the polygon into line segments. The homotopy in

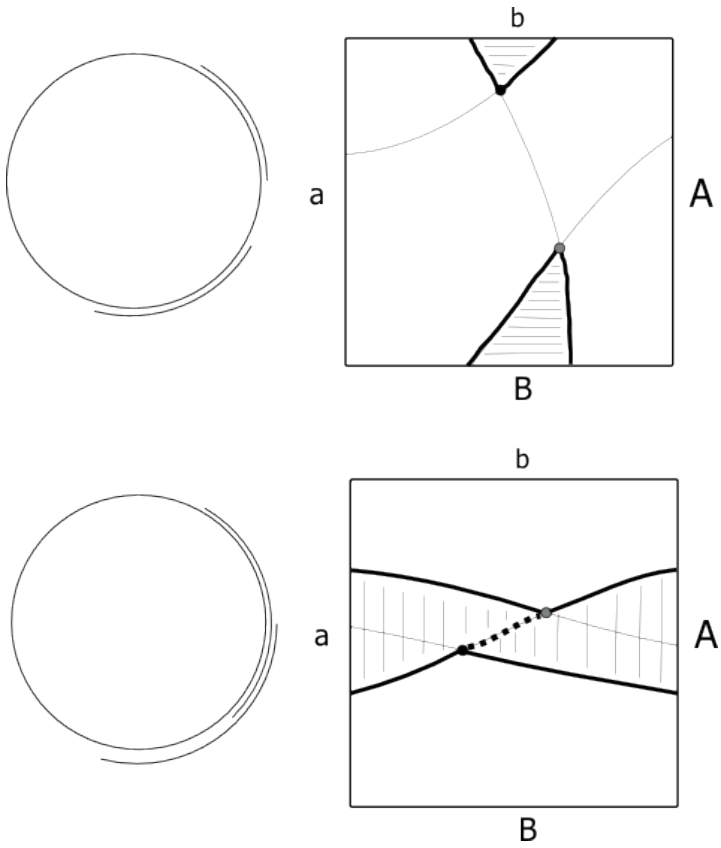


Figure 3.1: An example of a proper and improper bigon on the torus, along with schematic preimages on S^1.

Figure 3.2 which switches the locations of the white and grey points removes the bigon that is hidden in the figure. This procedure of identifying bigons and removing them via switching the locations of points along the boundary can be done entirely combinatorially, and is the central idea behind the first algorithm, which we call the *combinatorial homotopoy algorithm.*

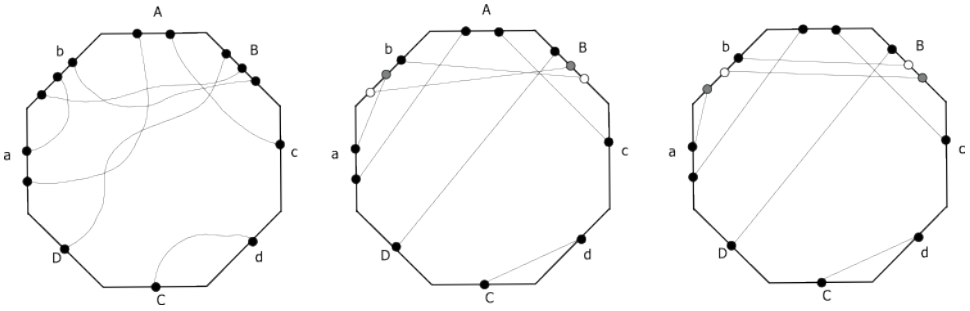


Figure 3.2: An arbitrary curve in $[a^2cdB^3]$ is first straightened out into line segments, and then modified to remove bigons.

3.1.2 The Combinatorial Homotopy Algorithm

The input for the algorithm introduced here will be the cyclic edge gluing pattern for a standard fundamental polygon (see Section 2.2.1 for the definition) as well as a cyclically reduced edge crossing sequence determining a free homotopoy class $[\gamma]$. The output will be a visual representation of the fundamental polygon and a curve in $[\gamma]$ which has the minimal possible number of self-intersections.

Definition 3.1.4. A *segmented representative* γ on a standard fundamental polygon P corresponding to a free homotopy class $[\gamma]$ with edge-crossing word w_γ is a curve γ in $[\gamma]$ such that:

- Each arc of γ crossing through P is a line segment with endpoints on

the interior of different edges of ∂P.

- All endpoints of these line segments are distinct.

Once such a γ is chosen, we can assign a label to each of the endpoints of these line segments using the labeling of the sides of P. We can then read these labels in clockwise order around ∂P to form a cyclic sequence, abbreviated P_γ. We can also form the cyclic sequence of *pairs* of labeled points which are connected by line segments, which we will abbreviate Q_γ.

Example 3.1.5. Using the segmented representative in the middle image of Figure 3.2, we could have $P_\gamma = a_1a_2b_1b_2b_3A_1A_2B_3B_2B_1c_1d_1C_1D_1$ and $Q_\gamma = (b_2a_2)(A_2a_a)(a_1c_1)(C_1d_1)(D_1B_3)(b_3B_1)(b_1B_2)$. In the third image of Figure 3.2, Q_γ remains the same, but $P_\gamma = a_1a_2\mathbf{b_2b_1}b_3A_1A_2B_3\mathbf{B_1B_2}c_1d_1C_1D_1$ in that case.

Proposition 3.1.6. *The number of intersections of a segmented representative γ is completely determined by P_γ and Q_γ.*

Proof. Any segmented representative determines the two lists, since the endpoints of all of the line segments are in general position by the second condition of our definition of segmented representatives. Now suppose that two segmented representatives have the same sequences P_γ and Q_γ (possibly after a relabeling). Since each arc passing through our polygon is a line segment, intersections are completely determined by the relative positions of the endpoints, which is precisely what P_γ and Q_γ record. □

Definition 3.1.7. Each pair of endpoints determining a line segment in the list Q_gamma will be called a *combinatorial segment*, abbreviated by some pair (p_n, p_m), where p_n and p_m are labels for the endpoints of the line segment.

Proposition 3.1.8. *A segmented representative corresponding to a cyclically reduced edge-crossing word W cannot contain a monogon.*

Proof. If some crossing x corresponds to a monogon, then we can find a sequence of directed Euclidean line segments starting and ending at x whose edge-crossing sequence is a trivial word. Since the fundamental group of our surface is a free group, there must be a pair of inverses adjacent to each other in this edge crossing sequence. That means there is a pair of inverses adjacent to each other in some cyclic permutation of W, contradicting the fact that W was reduced. □

So it remains to decide whether or not a given segmented representative contains a proper bigon. We have to check to see if there are intersections occuring in pairs which can cancel out with one another. At each intersection, there are four "sectors" which we could search in to find the paired intersection. If we choose an order for the two segments creating an intersection, then each of these four sectors is determined by a pair of orientations (+,+), (+,-), (-,+) or (-,-), since the two segments creating the intersection have a natural orientation.

Remark 3.1.9. On Notation: In what follows, if a combinatorial segment is abbreviated by some symbol $*$, then $*^1$ and $*^2$ will refer to the first and second points respectively in the combinatorial segment. We will also drop the adjective "combinatorial" when the context is clear.

Definition 3.1.10. A *combinatorial bigon with orientation (+,+)* is a pair of sequences of consecutive combinatorial segments in Q_γ, $W_i, W_{i+1}, ..., W_k$ and $W_j, W_{j+1}, ...W_l$ of equal size L, called the *length* of the bigon, so that:

1. W_{i+n} intersects $W_{j+n} \iff n = 0$ or $n = L - 1$

2. ${W_{i+n}}^2 = {W_{j+n}}^2 \ \forall n$ such that $0 < n < L - 1$

The two sequences described are called the *legs* of the bigon. We will abbreviate these two sequences of combinatorial segments as $W_{i...k}, W_{l...j}$

These conditions capture the intuitive idea of a pair of line segments intersecting, then fellow traveling without crossing for some time, and finally intersecting again. A similar definition can be given for combinatorial bigons with orientation $(+,-)$, with the indices adjusted to deal with the opposite orientation of the 2nd curve. By switching the roles of the initial and terminal intersection of a bigon, we may assume that any combinatorial bigon has orientation (+,+) or (+,-). Any bigon of a segmented representive in a punctured surface satisfies the conditions of the above definition, and so we have the following:

Proposition 3.1.11. *Bigons in a segmented representative are in one-to-one correspondence with combinatorial bigons*

Proof. It is clear that a combinatorial bigon determines a bigon. To go the other way, let us suppose that a segmented representative contains a bigon. Consider the universal cover of the surface, which can be realized as the unit disk. Since the surface is punctured, the disk can be tiled by ideal polygons with the standard gluing. Consider the collection of lifts of our curve to the universal cover. Since the curve contains a bigon, there will be two sequences of lifted segments which intersect in two distinct lifts of the fundamental polygon P_1 and P_2 (this is precisely condition 1) which bound a single disk. Since the polygons are ideal and we always use reduced cyclic words to construct segmented representatives, these two sequences of lifted segments must cross through the same sequence of edges of fundamental

polygons in order to start and end at P_1 and P_2. This is precisely condition 2. □

Corollary 3.1.12. *If a segmented representative has no combinatorial bigons, then it has the minimal number of self-intersections.*

Once we have a combinatorial bigon, we can permute the endpoints of the line segments in an attempt to mimic the homotopy which would eliminate the bigon. This is the so called *combinatorial homotopy* which gives the algorithm its name. This procedure does not always remove intersections, with the result depending on the nature of the combinatorial bigon. The next few arguments will analyze this situation and find the right conditions which guarantee that the procedure is successful.

Definition 3.1.13. A combinatorial bigon is *removable* if switching the positions of the endpoints of the "interior" segments of the combinatorial bigon removes the intersections to produce a new representative with two fewer intersections.

Theorem 3.1.14. *If the two combinatorial legs of a bigon $W_{k\ldots i}, W_{l\ldots j}$ contain no combinatorial segments in common, then the bigon is removable.*

Proof. We first note that each pair of intermediate segments $W_{k+1}W_{l\pm1}$,..., $W_{i-1}W_{j\pm1}$ becomes crossed, and then uncrossed as the permutations run their course, and in the end all of the intermediate combinatorial segments will be swapped (see Figure 3.3). This clearly contributes no additional intersections. Since the combinatorial segments in the combinatorial bigon are distinct, the permutations of the terminal pairs of points do not affect one another, so that W_k no longer intersects W_l, and likewise W_i no longer intersects W_j.

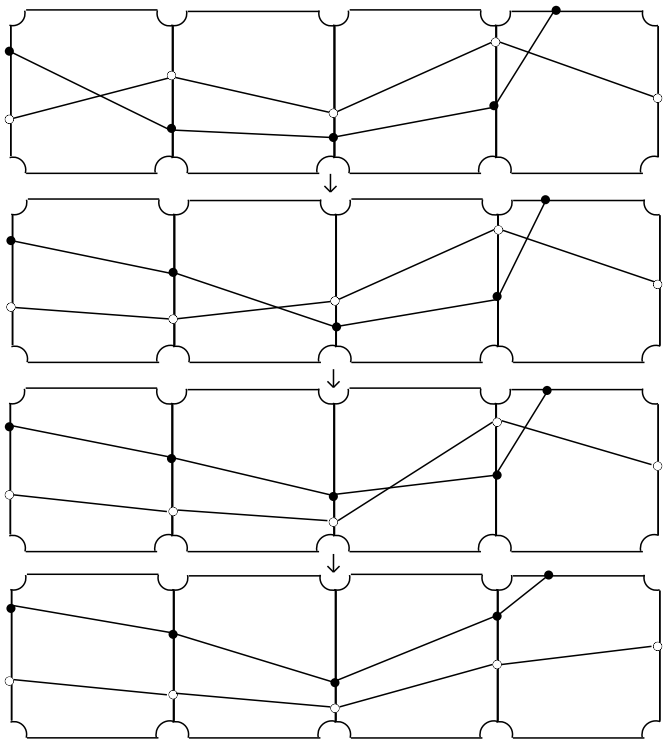


Figure 3.3: Geometrically, we are homotoping the bigon away. Combinatorially, we are switching the locations of the black and white points in P.

The only thing left to be checked is that no new intersections are formed during this process. To see this, consider Figure 3.4 where (a, b) and (c, d) represent the terminal segments of the bigon W_k and W_l. We assume our bigon has orientation $(+, +)$ since the argument is nearly identical for the $(+, -)$ case. This means that the final permutation to remove the bigon will swap the locations of the points labeled b and d.

Now suppose that there is another segment (x, y) which is fixed by the permutatation swapping b and d, which intersects (a, b) after permuting the appropriate points, but not before.

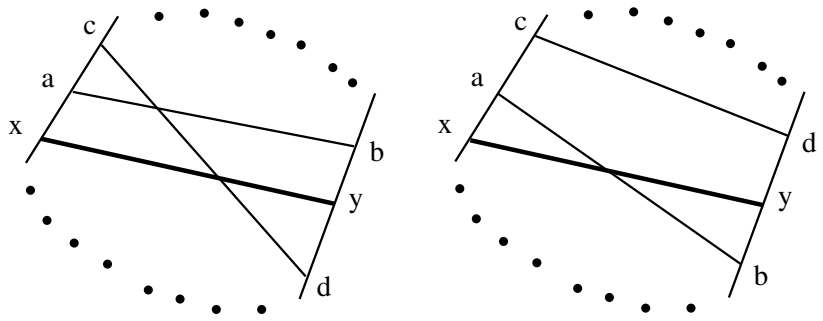


Figure 3.4: The net number of intersections decreases. The dotted curves represent the extra edges not involved in this particular collection of segments.

Since (x, y) did not intersect (a, b) before permuting b and d, the points a and b do not separate x from y along the boundary of the fundamental polygon. Since (x, y) intersects (a, b) after permuting b and d, the segment (c, d) must have intersected (x, y) to begin with. At once we see that after permuting b with d, (x, y) no longer intersects (c, d). Thus, for every intersection involving (a, b) that we create, we remove an intersection with (c, d). We arrive at a symmetric statement for intersections involving (c, d) and other combinatorial segments. A nearly identical argument works using the segments W_i and W_j, and so to avoid repetition we leave it to the curious. Combining the facts we have collected, we arrive at the desired result. □

Proposition 3.1.15. *Let $W_{k\ldots i}, W_{l\ldots j}$ be a combinatorial bigon.*

1. *If the combinatorial bigon has exactly 1 segment shared by both combinatorial legs, then it must be the case that $W_k = W_j$ or $W_i = W_l$*

2. *If there are exactly 2 segments shared by both combinatorial legs, then they must either be the first two segments of one combinatorial leg and*

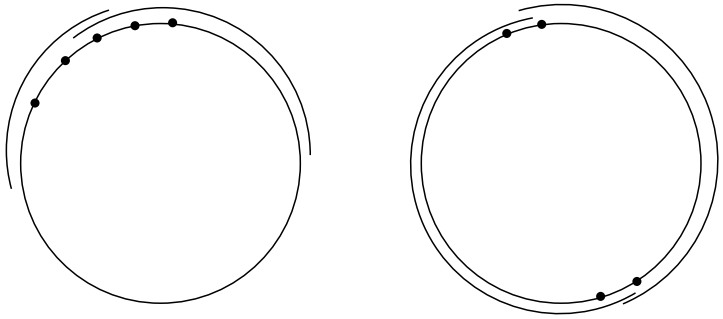


Figure 3.5: Two possible overlapping configurations. The points on the circle are preimages of the points where the segmented representative crosses an edge of the polygon. The arcs around the circle represent the preimages of the legs of the bigon.

the last two of the other, or the first and last of both.

3. *If there are more than 2 segments shared by both combinatorial legs, then the first two segments of one leg must be equal to the last two of the other leg.*

Proof. Let $f : S \to M$ be any choice of a segmented representative based on the point list P and segment list C that we have fixed for the moment. The preimages of the points where our representative crosses the edges of the polygon (i.e. the preimages of the points ${W_k}^1, {W_k}^2, {W_l}^1, {W_l}^2$, etc.) partition the circle into N arcs, where N is the length of the reduced cyclic word used to construct our representative. Each of the combinatorial legs can then be associated to N' of these arcs on the circle, where N' is the combinatorial length of the bigon. Let L_1 and L_2 be the arcs on the circle corresponding to the combinatorial legs. Since we are assuming that a segment is shared between the legs of the bigon, L_1 and L_2 must intersect. The two arcs

cannot be equal since the two legs of a bigon cannot be identical. It also cannot be the case that one arc is properly contained inside the other either, for then the two combinatorial legs could not have the same length. The only possibility is that the overlap must occur at the end of the arcs, which is exactly what statement 1 says. For statements 2 and 3 the argument is again very simple, except now there is the possibility the arcs L_1 and L_2 could wrap around and intersect on opposite sides of the circle. However, this situation is covered in the proposition. □

Theorem 3.1.16. *Let $W_{k\ldots i}, W_{l\ldots j}$ be a combinatorial bigon. The corresponding bigon in any segmented representative will be improper iff one of the following conditions holds:*

1. *At least two segments are shared by both combinatorial legs*
2. *One segment is shared by both combinatorial legs, say $W_k = W_j$, and ${W_l}^1$ is between ${W_i}^1$ and ${W_k}^1$, and ${W_i}^2$ is between ${W_l}^2$ and ${W_k}^2$ in the cyclic list of points P_γ.*

Proof. First choose a segmented representative and a fundamental polygon based on on P_γ and Q_γ. Let one leg of the geometric bigon be denoted L_1 and the other L_2. The preimages of all the points where our representative crosses the edges of the polygon(i.e. the preimages of the points ${W_k}^1, {W_k}^2, {W_l}^1, {W_l}^2$, etc.) partition the circle into N arcs, where N is the length of the reduced cyclic word used to construct our representative. Suppose that there are at least two segments shared by both combinatorial legs. This is equivalent to saying that the intersection $I = f^{-1}(L_1) \cap f^{-1}(L_2)$ is not entirely contained in a single partitioning arc. The only deformations of our segmented representative allowed are sliding the endpoints of segments

along the edges of the fundamental polygon without changing P. Any such homotopy of the line segment positions leaves the combinatorial description of the bigon unchanged. Therefore we still have that I is not contained within a single partitioning arc. With our construction, the endpoints of the preimages of L_1 and L_2 must be contained in the interiors of the dividing arcs - otherwise we would have two line segments emanating from the same point on an edge of the polygon. Since the endpoints of $f^{-1}(L_1)$ and $f^{-1}(L_2)$ must be contained in the interiors of dividing arcs, and since I is not contained within a single dividing arc, I cannot be empty. Figure 3.5 shows this situation.

This covers all cases except when exactly two segments are shared and that they occur at the ends of both combinatorial legs i.e. $W_k = W_j$ and $W_i = W_l$. From this we see that our bigon really only has one vertex, and thus that our bigon must be improper. For the 2nd condition above, our representative locally looks like Figure 3.6. The only way to remove the overlap is to translate one line segment over another, which clearly changes P. Thus our combinatorial condition guarantees that the bigon will be improper. □

Theorem 3.1.17. *Let $W_{k\ldots i}, W_{l\ldots j}$ be a combinatorial bigon with $W_k = W_j$, and no other segments shared. Then the bigon is removable if the 2nd condition from Theorem 3.1.16 is not satisfied.*

Proof. As in Theorem 3.1.14 we only need to consider the 6 points in the terminal segments, W_k,W_l, and W_i, since the intermediate segments are swapped. We consider all possible relative positions of those 6 points within P that simultaneously realize:

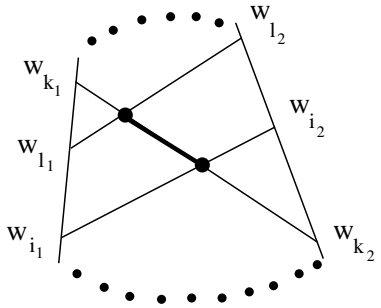


Figure 3.6: A configuration depicting an improper bigon. The dotted curves represent the extra edges not involved in this particular collection of combinatorial segments.

1. W_k intersects W_l
2. W_k intersects W_i
3. ${W_k}^2$ and ${W_l}^2$ on the same edge
4. ${W_k}^1$ and ${W_i}^1$ on the same edge.

These four conditions are necessary in order for our pair of sequences to actually be a combinatorial bigon. There are 12 configurations of points satisfying these conditions, and only 6 that need be considered once symmetry is taken into account. To see this, consider an oriented circle representing the cyclic ordering of P. We imagine placing the initial line segment, say W_k which then divides the circle into two arcs. Next, there are two choices for the placement of the line segment W_l, corresponding to which arc of the circle we wish to place the point ${W_l}^1$. For a given placement of W_l, the circle is divided into four arcs. The point ${W_i}^1$ may be placed on any arc except the one between the points ${W_k}^2$ and ${W_l}^2$ (that would imply ${W_i}^1$ is not on the same edge as ${W_k}^1$ or that ${W_k}^2$ and ${W_l}^2$ are on the same edge,

contrary to conditions 3 or 4. Once $W_i{}^1$ is placed, there are two choices for $W_i{}^2$ that result in W_k crossing W_i. Counting all possible combinations, we get $2 * 3 * 2 = 12$ total, but we can eliminate half of the sequences, since one ordering and its reverse ordering are equivalent for t he purpose of determining the effect of the permutations.

Below, the 6 essential orderings and total number of intersections they determine are given, along with their ordering and intersections after the permutations are performed:

1. $W_k{}^1$-$W_i{}^1$-$W_l{}^2$-$W_k{}^2$-$W_i{}^2$-$W_l{}^1$ **3** $W_i{}^1$-$W_k{}^1$-$W_k{}^2$-$W_l{}^2$-$W_i{}^2$-$W_l{}^1$ **1**
2. $W_k{}^1$-$W_i{}^1$-$W_l{}^2$-$W_k{}^2$-$W_l{}^1$-$W_i{}^2$ **2** $W_i{}^1$-$W_k{}^1$-$W_k{}^2$-$W_l{}^2$-$W_l{}^1$-$W_i{}^2$ **0**
3. $W_k{}^1$-$W_l{}^2$-$W_i{}^2$-$W_k{}^2$-$W_l{}^1$-$W_i{}^1$ **3** $W_i{}^1$-$W_k{}^2$-$W_i{}^2$-$W_l{}^2$-$W_l{}^1$-$W_k{}^1$ **1**
4. $W_k{}^1$-$W_i{}^2$-$W_l{}^2$-$W_k{}^2$-$W_l{}^1$-$W_i{}^1$ **2** $W_i{}^1$-$W_i{}^2$-$W_k{}^2$-$W_l{}^2$-$W_l{}^1$-$W_k{}^1$ **0**
5. $W_k{}^1$-$W_i{}^2$-$W_l{}^2$-$W_k{}^2$-$W_i{}^1$-$W_l{}^1$ **3** $W_i{}^1$-$W_i{}^2$-$W_k{}^2$-$W_l{}^2$-$W_k{}^1$-$W_l{}^1$ **1**
6. $W_k{}^1$-$W_l{}^2$-$W_i{}^2$-$W_k{}^2$-$W_i{}^1$-$W_l{}^1$ **2** $W_i{}^1$-$W_k{}^2$-$W_i{}^2$-$W_l{}^2$-$W_k{}^1$-$W_l{}^1$ **2**

Case 6 satisfies the second condition in Theorem 3.1.16 and so determines an improper bigon, so it is not surprising that the permutations do not reduce intersections. The only thing that remains to be proven is that for each of the 5 "good" cases, no additional intersections are produced. To make things easier to keep track of, let us relabel the segments as follows: $W_k = (a, b)$,$W_l = (c, d)$, and $W_i = (e, f)$. We'll prove the theorem for the first case above by following a line of reasoning similar to that in Theorem 3.1.14. Careful consideration of Figure 3.7 yields the proof, but we give some of the details here. Suppose a segment $(x, y) \neq (a, b), (c, d)$ nor (e, f), does not intersect (a, b) initially, but does after permuting. Then (x, y) must have

intersected one of (e, f) or (c, d) beforehand. To see this, note that one of the points x or y must lie in the "top" arc between e and d, while the other point must have been either between a and e, or d and b. However, after the permutations, it cannot intersect either of them, as seen from the figure and the argument in the previous sentence. Now suppose (x, y) intersects (c, d) after permuting but not before. Since the point c is unchanged by a permutation, we see that (x, y) must have intersected both (a, b) and (e, f) before the permutations. But after the permutations, (x, y) could not possibly intersect (a, b), since then (x, y) would have intersected (c, d) to begin with. Finally, suppose (x, y) intersects (e, f) after the permutations but not before. The point f is fixed by the permutations, and we see that (x, y) must have intersected both (a, b) and (c, d) before the permutations. By the exact same reasoning as in the previous situation, (x, y) can no longer intersect (a, b) after the permutations.

In every situation, if an intersection is introduced, there is a corresponding intersection that is removed. Thus there is no net gain of intersections between (x, y) and the segments $(a, b), (c, d)$ and (e, f). Since the segment (x, y) was arbitrary, we proved the theorem for case 1. The proofs for the other 4 "good" cases are nearly identical, so we leave them to the curious. □

Corollary 3.1.18. *Every proper bigon in a segmented representative corresponds to a removable combinatorial bigon.*

Remark 3.1.19. In Hass and Scott, it is proven that if a representative does not have the minimal number of intersections, then it has a *proper* bigon. The notions of removable and proper do not actually coincide, but since every proper bigon is removable there is no logical contradiction.

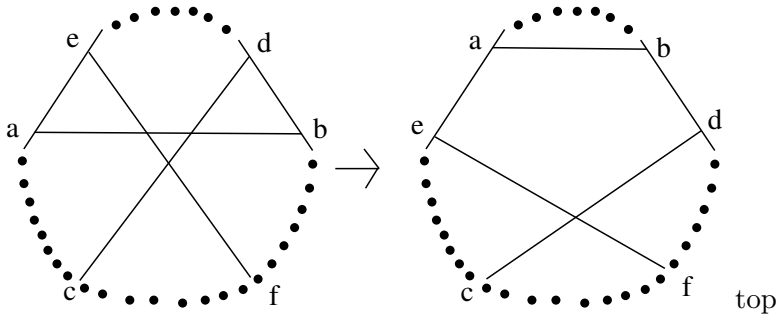


Figure 3.7: The new configuration after permuting the appropriate points. The dotted curves represent the irrelevant edges of the fundamental polygon.

3.1.3 The Minimal Linking Algorithm

The previous algorithm essentially has as its input the lists P_γ and Q_γ which record the cyclic order of points along the boundary, and the line segments making up the representative (since it takes the surface word X and the curve word W and arbitrarily creates P_γ and Q_γ). It then modifies P_γ in order to produce a representative with minimal intersection number. The following algorithm also takes in X and W and produces a P_γ and Q_γ which correspond to a representative with the minimal intersection number. The difference is that the next algorithm does not analyze these cyclic lists, but instead produces them in a way which guarantees minimality. Hence it is less flexible, but more efficient.

Imagine tiling the hyperbolic plane with fundamental polygons, giving these polygons edge labels matching those of your surface word X in a clockwise fashion. Each geodesic in $\mathbb{H}^2$ can then be associated with (almost) canonical bi-infinite word with a marked *center* describing the edge-crossing sequence in both directions from a central chosen fundamental polygon.

There is only ambiguity if the geodesic crosses through a vertex of a polygon, which cannot happen in the case of a punctured surface with a standard fundamental polygon, since all vertices of the polygon will be at infinity. Closed geodesics will correspond to words which are periodic, since the edge crossing sequence must repeat itself if the geodesic is to end up where it started.

For the following arguments, we will assume that we have a fixed standard polygon, and a fixed cyclically reduced word $W = w_1 w_2 ... w_n$. corresponding the edge crossing sequence of some closed geodesic curve γ. We first introduce some definitions, which are a modification of the ideas in [10],[13] and [16], which relate geometric and combinatorial properties of a geodesic and its associated edge-crossing word.

Definition 3.1.20. Two geodesics are *hyperbolically linked* if their endpoints alternate on $\partial\mathbb{H}^2$.

Definition 3.1.21. Two subwords of W of length one, w_j and w_k, are said to be a *short link* if the sequence of letters w_{j-1}^{-1},w_{k-1}^{-1},w_j,w_k has no repetitions and appears in clockwise order in the labeling of the fundamental polygon.

Definition 3.1.22. Two subwords of W of length $l > 1$, $W_j = w_j w_{j+1} ... w_{j+l-1}$ and $W_k = w_k w_{k+1} ... w_{k+l-1}$ are said to be a *parallel long link* if:

1. $w_{j+i} = w_{k+i} \forall\ 0 < i < l-1$
2. $w_j^{-1} \neq w_k$ and $w_{j+l-1} \neq w_{k+l-1}$
3. The two sequences of letters w_j^{-1},w_k^{-1},$w_j = w_k$ and w_{j+l-1},w_{k+l-1},$w_{j+l-2} = w_{k+l-2}$ appear in clockwise order in the labeling of the fundamental polygon.

Definition 3.1.23. Two subwords of W of length $l > 1$,$W_j = w_j w_{j+1} ... w_{j+l-1}$ and $W_k = w_k w_{k+1} ... w_{k+l-1}$ are said to be an *alternating long link* if W_j and ${W_k}^{-1}$ form a parallel long link.

A short link intuitively corresponds to a pair of arcs crossing in the fundamental polygon after emerging from distinct sides. A parallel long link corresponds to a pair of arcs which enter the same sequence of edges together for some time before finally crossing, and an alternating long link is similar except that one of the arcs has the opposite orientation.

A word W in a group is said to be *primitive* if W cannot be expressed as a power of another word (besides W^{-1}). Non-primitive edge-crossing sequences correspond to curves which essentially run parallel to some core curve a number of times. The geodesic representative of such a curve coincides entirely with the geodesic representative of the underlying core curve, so arguments about transverse intersection numbers need to be tweaked. We will deal with this special case in the next few statements.

Proposition 3.1.24. *If W is a primitive word corresponding to a closed curve with k self-intersections, then W^n has $n^2k + n - 1$ self-intersections $\forall n \in \mathbb{N}$.*

Proof. Arrange n copies of the core curve in parallel to one another on the surface. For each of the intersections of the core curve, we obtain a local grid-like pattern, producing n^2 intersections. Now, connect the n strands as shown in Figure 3.8 to obtain a curve representing W^n with $n^2k + n - 1$ self-intersections. None of the intersections coming from one of the grids can be part of a proper bigon, since by our assumption that W is primitive, the intersecting arcs must eventually exit through different sides of the polygon. The intersections formed by connecting the parallel strands are all part of

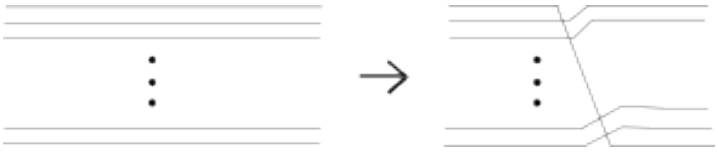


Figure 3.8: Arrange n copies of a simple closed curve in parallel. Connect them as shown to obtain a curve with n-1 intersections.

improper bigons, so our curve with $n^2k + n - 1$ intersections does achieve the minimal number of transverse self-intersections. □

Now that we have disposed of the non-primitive case, we can turn to the general statement:

Theorem 3.1.25. *If W is primitive, then every intersection in a minimal representative of W corresponds to a unique short, parallel, or alternating link.*

Proof. First, we prove that each intersection corresponds to at least one link. This amounts to showing that starting from an intersection, the arcs emanating from the point of intersection must eventually leave through different sides of the fundamental polygon. If W is primitive, they cannot fellow travel in perpetuity. Let x be a point of intersection. Suppose that as we follow the pair of arcs determining this intersection, the arcs always cross the same sequence of edges. The initial arcs determining the intersection correspond to different locations in the edge-crossing sequence W, say w_j and w_k. Eventually, we follow the arcs long enough so that the arc starting on w_j reaches the arc w_k, and the pattern repeats itself. Thus W is periodic, a contradiction. Now, we prove uniqueness. There are four pairs of edge labels associated to an intersection x, which can be arranged cyclically, corresponding to the four "sectors" determined by this intersection.

For two of these sectors, the orientations of the arcs determining the sector are identical, and for the other other two, the orientations are opposite one another. At most two of these pairs have edge labels which are identical, since adjacent pairs cannot both have identical edge labels (otherwise our word would not be reduced). If none of the pairs of edge labels contain identical labels, then x corresponds to a short link, and precisely one ordered pairing of the arcs gives the clockwise pattern required in the definition of a short link. Suppose then, that some pair of edge labels are identical. By our earlier argument, if we follow these arcs, we know they must eventually exit through different sides of the polygon. If the orientation of both arcs is the same, then the opposite sector also has the same orientation. Following the arcs in both sectors until they leave different edges of the fundamental polygon determines a parallel long link. Only one ordered pairing of these two sequences of paths gives the clockwise pattern required in the definition of a parallel long link. A similar argument holds for an alternating long link. Now, if an intersection is associated to a parallel long link, it cannot also be associated to an alternating long link, because only the opposite sectors determined by an intersection can have matching edge labels. □

The previous theorem is essentially a restatement of the main theorem in [16], but proved in a new way and with new combinatorial objects:

Theorem 3.1.26. *(Cohen-Lustig[16]) Intersections in a minimal representative of a primitive free homtopy class with word W are in bijection with the set of short, parallel, and alternating links in W.*

Each letter in our word W corresponds to a segment crossing through our fundamental polygon. Each link in W then corresponds to a sequence of pairs of segments that emerge from distinct edges, which then possibly fellow

travel through identically labeled edges for some time, and then pass through different edges in a way that creates an intersection. In the arguments that follow, we will blur the distinction between a long link, and the sequence of segments associated with it.

Definition 3.1.27. Let us say that a link has been *positioned* if it has been determined which pair of segments cross to determine the intersection

Constructing a representative with minimal intersect amounts to positioning all of the links in a coherent way so that no additional crossings are introduced.

Algorithm 3.1.28. The following algorithm with input a standard surface word X and a cyclically reduced curve word W produces a representative with the minimal number of intersections:

1. Form the list of links uniquely determined by X and W.
2. Choose a side of the polygon and consider the collection of arcs entering the same edge of the polygon.
3. Choose a pair of arcs along this side and ask following: Are these arcs part of a link?
    - If not, then you know their relative positions are such that they do not cross.
    - If so, let the two arcs under consideration be the crossing point, which determines their relative positions along the boundary.
4. Until all pairs have been exhausted, choose a new arc and pair it with all of the arcs that have been considered up to this point, asking the question in step 3.

5. Repeat steps 2 through 4 for a new side that is not the inverse of a previously considered side until all remaining sides have been exhausted.

Proof. Clearly, this algorithm produces some representative of our free homotopy class. We must however, demonstrate that at each step of this process, no bigons are formed so that the final representative has the minimal number of intersections. The algorithm as described determines the relative positions of the arcs entering a side of the fundamental polygon. By Proposition 3.1.6 we know that this information determines the topological structure of the representative, and thus determines the presence of bigons. If there is a bigon, then at some step of the algorithms execution, we must have introduced an excess crossing which does not correspond to a link. But the only time when crossings are introduced is in Step 3, which occurs precisely when two arcs are part of a link. Since the algorithm checks all previously analyzed arcs at each step, it ensures that any subsequent arcs are properly placed with respect to all previously positioned arcs. □

Remark 3.1.29. This algorithm does not produce a canonical representative. Depending on the starting side, different configurations may be realized at the conclusion of the algorithm.

3.1.4 Examples

We will now show how the first algorithm works for the surface word **abAB** and the reduced cyclic word **bbAAA**, the same surface and curve as in Figure **??**. Recall that in Figure **??**, we have chosen $P_\gamma = a_1, a_2, a_3, b_1, b_2, A_3, A_2, A_1, B_2, B_1$ and $Q_\gamma = (B_1, A_3)(a_3, A_2)(a_2, A_1)(a_1, b_2)(B_2, b_1)$ as our initial representative. We now check through pairs of segments until we find a pair that intersect. Upon inspection, we see that $(a_3, A_2)I(B_2, b_1)$. The next step

is to see if this vertex can possibly be part of a bigon. There can be no bigon starting at this vertex with a $(+,+)$ orientation, since the segments spit to the A and b sides. Likewise for the other 3 orientations. We conclude that this particular intersection cannot be a vertex of a bigon and move on until we find another pair of segments that intersect. Suppose the next pair of segments we find to intersect are (B_2, b_1) and (a_1, b_2). For this particular intersection, we see that it may be a vertex of a bigon with a $(+,+)$ orientation, since the points b_1 and b_2 are on the same edge. The next pair of segments we compare are (B_1, A_3) and (B_2, b_1), which also intersect. Thus we have a combinatorial bigon $\{(B_2, b_1)(B_1, A_3); (a_1, b_2), (B_2, b_1)\}$. Since (B_2, b_1) is shared by both sequences, we must use Theorem 3.1.17 to check if we can remove this bigon to reduce the number of self-intersections. We see that this particular bigon is of type (5) in reverse, and so we may proceed. We need to switch b_1 with b_2 and B_1 with B_2, as shown in Figure 3.9.

Now, we have $P_\gamma = a_1, a_2, a_3, b_2, b_1, A_3, A_2, A_1, B_1, B_2$, while Q_γ remains the same. We must again compare pairs of segments until we find an intersection. We see that $(a_1, b_2) I (a_2, A_1)$, so we check to see if this intersection can be a vertex of a bigon. A $(-,-)$ orientation is ruled out, since b_2 and A_1 are on different edges, but we see that we may check for a $(+,+)$ orientation. Doing so results in the combinatorial bigon:

$$\{(a_1, b_2), (a_2, A_1), (a_3, A_2); (a_2, A_1), (a_3, A_2), (B_1, A_3)\}.$$

This bigon has two segments shared by both sequences, so it is guaranteed to be improper by Theorem 3.1.16. By Theorem 4.1.3 there must be a proper bigon if our representative does not have minimal self-intersection, so we continue to look for a different bigon to potentially remove. Sup-

pose we next find that $(a_1, b_2)I(a_3, A_2)$. We find that it is the vertex of the bigon $\{(a_1, b_2), (a_2, A_1); (a_3, A_2), (B_1, A_3)\}$. We switch a_1 with a_3, and A_1 with A_3, as shown in the second step of Figure 3.9. Finally, we once again check all pairs of segments using the permuted P_γ, and determine that there are no more proper bigons, and thus by Theorem 4.1.3 our representative has the minimal number of self-intersections possible. The final output of the algorithm is P = $a_3, a_2, a_1, b_2, b_1, A_1, A_2, A_3, B_1, B_2$ and C = $(B_1, A_3)(a_3, A_2)(a_2, A_1)(a_1, b_2)(B_2, b_1)$, which determines our representative.

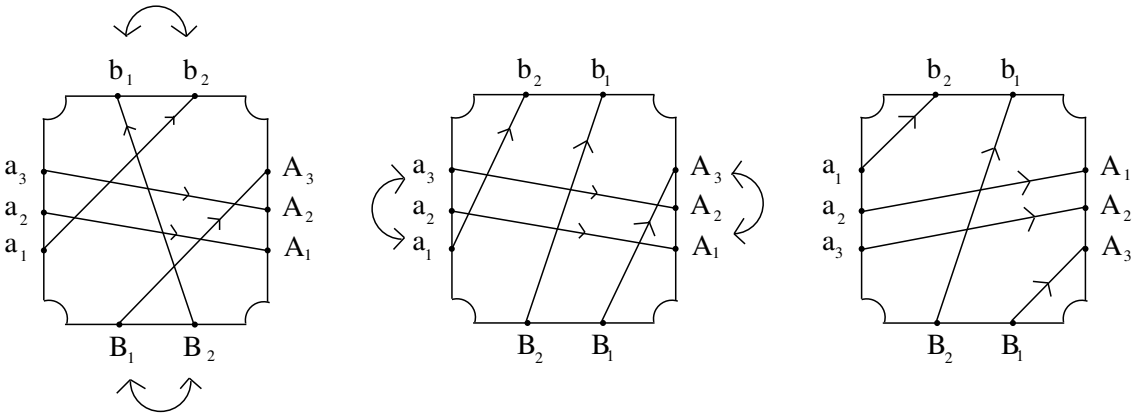


Figure 3.9: Sequence of permutations removing all of the bigons.

We now illustrate the second algorithm using the same example surface and curve. Let w_i be the i^{th} letter in $bbAAA$. Then upon inspection, we see that the pairs (w_4, w_2) and (w_5, w_2) form the only links, and both of them are short. Iterating first through the three arcs entering the A side of the polygon, we see they must be arranged without crossing, since the only links occur with an arc going from B to b. Similarly for the arcs entering the b edge, which are not linked with each other. This forces the configuration to

be the same as the rightmost image in Figure **??**.

3.2 Filling Curves

Definition 3.2.1. A collection of curves $\{\gamma_i\}$ is said to be *filling* if $S_g - \{\gamma_i\}$ is a disjoint union of topological disks.

Collections of filling curves play important roles in the study of Teichm´uller space and the mapping class group. For example, a pair of simple closed curves which are filling determines a pseudo-Anosov element of the mapping class group, and the dynamics of such a map are related to the combinatorial structure of the two simple curves (see [36]). Every collection of filling curves has a unique point in Teichm´uller space which minimizes the curves in the collection (see [26]). It is thus natural to ask for some procedure to determine whether a given collection of free homotopy classes is filling. We will show that the question of determining whether a collection of curves is filling can be answered by the algebraic data of the words representing each curve. There is a natural "greedy" algorithm to determine this, which we now describe, but it is computationally ineffective and not geometrically elegant.

Proposition 3.2.2. *Let W_i be a cyclically reduced multiword in $\pi_1(\S_g)$ corresponding to a collection of free homotopy classes of total length L. Suppose further that the collection of curves has geometric intersection number K. If $\{W_i\}$ is not filling, then there is a curve whose word length is less than some constant $C(L, K)$ which does not intersect a curve in W_I.*

Proof. If $\{W_i\}$ is not filling, consider a closed curve δ that traces out a portion of a regular neighborhood of the geodesic realization of $\{W_i\}$ on $\S_g$.

An upper bound for the word length of δ can be obtained by noting that each intersection of $\{W_i\}$ contributes at most 4 to the word length of the regular neighborhood boundary curve, and that δ can traverse portions of each arc of $\{W_i\}$ in a fundamental polygon P at most twice. $\square$

The above proposition then provides a brute force method to determine whether a collection of curves is filling: just check all words of length less than $C(L, K)$ for intersection. If there is a nontrivial word which has zero geometric intersection with W_i, then W_i is obviously not filling.

3.3 A new detection method

Definition 3.3.1. The *essential subsurface* of a collection of curves $\gamma_i \subset S$ is the smallest complexity π_1-injective subsurface S' which contains γ_i.

Definition 3.3.2. The *relative boundary* of an essential subsurface S' is the collection of words corresponding to the boundary curves of the essential subsurface.

When there is no danger of ambiguity, we will often refer to the relative boundary of a collection of curves γ_i, meaning the relative boundary of the essential subsurface corresponding to γ_i. A collection of curves on S is filling if and only if its essential subsurface is S itself, or equivalently that the relative boundary consists of trivial words. The relative boundary of any collection of curves γ_i can be computed in terms of the words corresponding to the γ_i, as the following arguments will show.

If we imagine that a collection of curves γ_i is presented to us as a system of intersecting arcs within a fundamental polygon, there is a natural intuitive procedure to determine whether or not the curves are filling. We simply trace

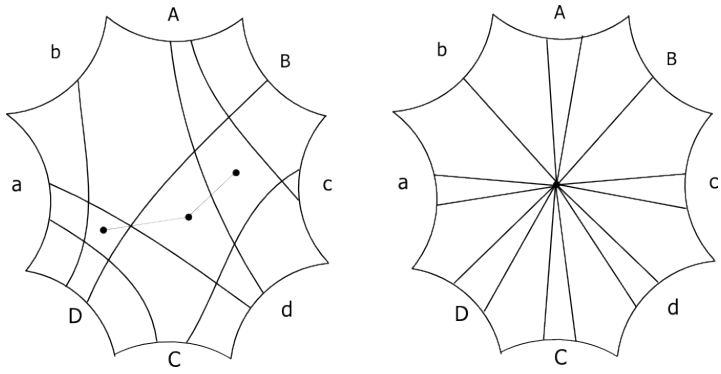


Figure 3.10: The configuration on the right is obtained by compressing all of the interior disks to points. The collection of curves on the left is filling only if the homotoped collection on the right is filling.

out along the arcs, with the convention that we "turn right" at each point of intersection. When we hit an edge of the polygon, we move over to the identified edge and continue tracing out the arc, with the same convention. Eventually, we will come back to the place where we started, and the edges of the polygon that we crossed in tracing out the arcs will spell out a word w. If this word is not trivial, then it represents part of the relative boundary of γ_i. The goal then, is to figure out an appropriate combinatorial method which is purely a function of the edge-crossing words $\{w_{\gamma_i}\}$ corresponding to $\{\gamma_i\}$. In order to do this, we need a method for detecting intersections which mimics the tracing procedure. From either the algorithms in Section 3.1, we can obtain a segmented representative of a free homotopy class which realizes the minimal intersection number. We will use the combinatorics obtained from the output of these algorithms to mimic the above intuitive procedure. See Figure 3.10 for the main motivating idea.

Definition 3.3.3. A *maximal linked chain* of segments W_1,W_2,...,W_k starting at an endpoint w of W_1 along ∂P is a sequence of segments such that:

1. Each W_i crosses each W_{i+1}
2. Each $W_j, j > 1$ has an endpoint which is cyclically closer to w in the clockwise direction than any other segment intersecting W_{j-1}
3. There is no segment intersecting W_k which has an endpoint closer to w in the clockwise direction than the endpoint of W_k closer to w in the clockwise direction.

. This closest endpoint of W_k will be called the *terminal point of the chain* and the edge the terminal point lies on will be called the *terminal edge*.

The definition of maximal linked chain captures the idea of following the curve and turning right at intersections until you hit a boundary edge of the polygon.

Definition 3.3.4. A *combinatorial boundary* is a cyclic collection of maximal linked chains $C_1, C_2, ..., C_m$ such that the starting point of C_{i+1} is identified with the terminal point of C_i $\forall 0 \geq i < m$ (treat indices modulo m). The *combinatorial boundary word* is the cyclic sequence of terminal edges for $C_1, ..., C_m$.

Theorem 3.3.5. *Every non-trivial word in the relative boundary has a non-trivial combinatorial boundary word associated to it.*

Proof. The argument is identical whether we use a single curve or multi-curves, so assume that some curve γ with word w_γ has a non-trivial relative

boundary. This means that if we take a small regular neighborhood of the geodesic representative of γ, the resulting surface with boundary will have a boundary curve which is not null-homotopic. The edge-crossing sequence of this curve will thus be a word which cannot be cyclically reduced. We will demonstrate that the edge-crossing sequence of this curve is precisely the sequence of terminal edges of combinatorial boundary. Since our boundary curve is not trivial, it crosses an edge somewhere. Now, follow this little piece of the boundary curve emanating from an edge. Since it is part of the relative boundary of γ, it cannot intersect γ, and so it must run parallel to various pieces of γ, turning right at each intersection. After a certain number of turns, it will exit again through another point along the edge of the polygon. It is impossible to determine the sequence of turns with only combinatorial information, since configurations can vary by triangle moves even with the same combinatorial data. However, we will show that a maximal linked chain can tell us where this boundary curve must exit the polygon. Suppose that the piece of the boundary b is initially running parallel to some segment W_1, and suppose that W' is any segment that intersects W_1. Then the point at which b exits the polygon must between the endpoint of W_1 and an endpoint of W' (the correct endpoint to consider obviously depends on the orientation of the segments W_1 and W_2). Take W_2 to be the line segment intersecting W_1 which has an endpoint closest to the point b_0 where b emerged from the edge of the polygon. If W_2 intersects another segment W'' with an endpoint closer to b_0, then again, the point where b exists the polygon must lie between the appropriate endpoint of W'' and b_0, since b will have been forced to turn at the intersection of W_2 and W'' if there were no other line segments to consider. Constructing this sequence of line segments is precisely constructing a maximal linked chain.

□

Thus, the algorithm to detect whether or not a word w_γ represents a filling curve can be summarized as follows:

Algorithm 3.3.6. Given a surface word X and a cyclically reduced word w_γ, the following algorithm will determine if γ is filling:

1. Determine the relative positions of all of the arcs in γ using either of the algorithms described in Section 3.1.

2. Use these relative positions to determine all of the maximal linked chains.

3. Combine the linked chains to form the combinatorial boundary words.

4. If all of the combinatorial boundary words are trivial, then γ is a filling curve on the surface determined by X.

3.3.1 An example

Consider the curve $AdbCBB$ on the surface of genus 2. In Figure **??** we show a minimal configuration and the sequence of linked chains that yield the boundary word, which can be reduced to the trivial word. Thus, $AdbCBB$ is a filling curve.

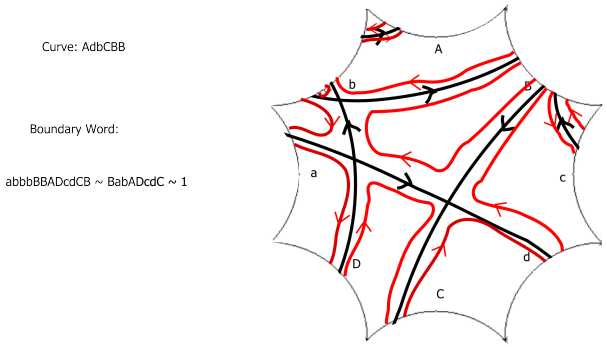


Figure 3.11: The arcs in red correspond to linked chains, which combine to form the boundary word $abbbBBADcdCB$, which can be reduced to the trivial word.

Chapter 4

Constructing filling curves

The combinatorial algorithms described in the previous chapter allow for experimentation and statistical analysis of properties of curves. See [14] for some results in this vein. This chapter will deal with results inspired by experimenting with the previously described algorithms.

4.1 Topological considerations

The first question we must ask is: What is the minimum possible number of self-intersections for a filling curve? The following proposition provides a lower bound for this topological question.

Proposition 4.1.1. *If a curve γ fills a surface S_g, then γ has at least $2g-1$ self-intersections.*

Proof. A curve with k self-intersections whose complement consists of d disks yields a cell decomposition of the surface with k vertices, 2k edges, and d faces. Now observe that $\chi(S_g) = 2 - 2g = d - k$, so k scales linearly with d, being minimal when $d = 1$. □

Examining the proof of the proposition, we have the following more general statement:

Corollary 4.1.2. *If a collection of curves $\{\gamma_i\}$ fills a surface S_g with d complementary regions, then $\{\gamma_i\}$ has $2g-2+d$ total intersections.*

Now that we have a lower bound, we can ask whether it is actually realized for some curve, or collection of curves. For a single curve γ which has self-intersections, this minimum is realized, which is the content of Theorem 4.3.1. Interestingly enough, when $g = 2$, there are no *pairs* of simple curves that realize the minimum of 3 intersections. See [4].

First, we need a criterion to determine when a collection of curves on a surface has the minimal possible number of intersections among all possible sets of curves in the same collection of free homotopy classes. This is an important condition, because an arbitrary curve can be turned into a filling curve by sufficiently distorting it on the surface. If a surface has a hyperbolic metric, which is the case we'll be considering, then each free homotopy class has a unique geodesic, and these geodesics always minimize the intersection number. Recall the theorem of Hass and Scott which we used earlier:

Theorem 4.1.3. *([20]) A map $f : S^1 \to S_g$ has the minimal possible number of self-intersections if f has no monogons or proper bigons.*

The strategy then, will be to produce topological configurations of curves which have no monogons or bigons. The geodesic in an associated free homotopy class will have the same number of intersections, and roughly equivalent combinatorics (this subtly will be discussed in more detail later).

4.2 Combinatorial tools

We now introduce some combinatorial tools which will be the main ingredients in constructing filling curves.

Definition 4.2.1. A *ribbon graph* (or *fat graph*) is a graph Γ with a chosen cyclic ordering of the half-edges at each vertex of Γ.

If we take a collection of curves on a surface whose union is connected, then the collection of curves determines a 4-valent ribbon graph, since the surface determines a natural cyclic ordering at each point of intersection. This process can be reversed:

Lemma 4.2.2. *A 4-valent ribbon graph Γ determines a multi-curve γ on a surface with boundary $S(\Gamma)$.*

Proof. Embed Γ in $\mathbb{R}^3$ and use the cyclic ordering at each vertex to form a 2-dimensional regular neighborhood of the graph locally at each vertex. Then extend this regular neighborhood along the edges of Γ in a way that preserves orientation. The result of this process will be an orientable surface with boundary. □

Remark 4.2.3. Once a "starting point" and a direction for each component of the multi-curve γ is chosen, there is a canonical way to orient each intersection.

We will use the phrase "ribbon graph" in place of "surface with boundary determined by a ribbon graph" to improve the exposition.

Definition 4.2.4. A ribbon graph Γ is said to be *minimal* if, after attaching disks to the boundary of the ribbon graph, the graph corresponds to a

configuration with the minimal possible number of intersections in its free homotopy class.

If a ribbon graph is not minimal, then the associated curve has a monogon or a bigon, by Theorem 4.1.3. It is easy to construct ribbon graphs which are not minimal by creating monogons or bigons. Determining whether a given ribbon graph is minimal or not is more delicate, and not always obvious simply by looking at the graph Γ.

Lemma 4.2.5. *If a 4-valent ribbon graph Γ is not minimal and the defined curve contains a monogon, then there is a vertex v and an oriented smooth path p of edges e_0,e_1,...,e_n starting and ending at v such that:*

- *Every path of edges p' starting at an intersection with p with orientation o, also intersects p at another point with orientation $-o$.*
- *If p' is a path as in the previous item, then if p'' is any path intersecting p', then p'' also intersects p.*

Proof. We will unwind the combinatorial conditions in the lemma and show that a monogon satisfies the properties of the path p in the lemma. Let us recall the definition of a monogon: If we identify the curve as a map $f : S^1 \to S_\Gamma$, then f has a monogon if there is an arc α in S^1 so that f identifies the endpoints of α and $f|\alpha$ is nullhomotopic. In the lemma, α is just the path corresponding to the edges $e_0, ..., e_n$. The endpoints are identified since the path starts and ends at the vertex v. If this path is nullhomotopic, then it bounds a (possibly non-embedded) disk D. Every path that enters D, which we will call an *interior path*, must eventually exit D with the opposite orientation it entered with. This is precisely the second condition. Finally, we know that if any path intersects one of the interior

paths, it must have entered the disk from somewhere, which is the third condition. This proves the necessity of the conditions. □

Now let us assume that a curve has a bigon, and see what combinatorial properties must be present. Notice that when a bigon is present, both of the arcs of the bigon have a well-defined "inner" side and "outer" side.

Lemma 4.2.6. *If a 4-valent ribbon graph Γ is not minimal and the defined curve contains a bigon, then Γ has a pair of vertices v and w connected by paths p_1 and p_2 such that:*

- *Each path p' intersecting p_1 or p_2, must also intersect p_1 or p_2, with the appropriate orientations at the points of intersection.*
- *If p' is a path as in the previous item, then if p'' is any path intersecting p', then p'' also intersections p.*

Proof. Again, we must unravel the combinatorial conditions in the lemma and show that they imply that a curve with a bigon has the stated properties. The proof is only a slight modification of the previous proof. Let us recall the definition of a bigon: If we identify the curve as a map $f : S^1 \to S_\Gamma$, then f has a bigon if there are disjoint subarcs α and β in S^1 so that f identifies the endpoints of α and β and $f|\alpha \cup \beta$ is nullhomotopic. In the lemma, α and β are just the paths corresponding to p_1 and p_2. The endpoints are identified since the paths have endpoints v and w. If this concatenation of paths is nullhomotopic, then it bounds a (possibly non-embedded) disk D. Every path that enters D from the well defined "outer side", must eventually exit D through the "inner side". This is what the second condition is saying. And again, we know that if any path intersects one of the interior paths, it must have entered the disk from somewhere, which is the third condition. □

In Figure 4.1, we see that the configuration on the left fails the second property, while the right figure satisfies all of the conditions to form a bigon, which is clear in the image.

Figure 4.1: The left configuration does not determine a bigon, while the right one does.

4.3 Constructions

If we have a curve γ that fills S_g and whose complement is a single disk, then if we take a small enough regular neighborhood of the curve and delete the disk, we end up with a topological surface of genus g with one boundary component, which defines a ribbon surface. In this case, γ itself is the 4-valent graph, which naturally has a cyclic ordering at each vertex using the intersecting arcs of γ. Conversely, if we start with a topological surface of genus g with one boundary component, constructed in such a way that it is a regular neighborhood of a 4-valent graph with a cyclic ordering at each vertex, then γ represents a filling curve whose complement is a single disk. Thus we can use a ribbon surface to implicitly give us γ. Note that the curve γ constructed this way is only defined up to homeomorphism and does not

correspond to a specific free homotopy class. We also note that for a given underying graph Γ, there are many choices (up to ambient isotopy in $\mathbb{R}^3$) for choosing a ribbon surface with Γ as the core graph. And now, onto the construction of a self-intersecting curve with the minimal possible number of self-intersections:

Theorem 4.3.1. *For $g \geq 2$, there exists a curve γ with $2g-1$ self-intersections, whose complement is a single topological disk.*

Proof. We will construct a particular ribbon surface when g=2 realizing the minimum of 3 intersections and then prove by induction that this can be modified to produce a ribbon surface for arbitrary higher genus.

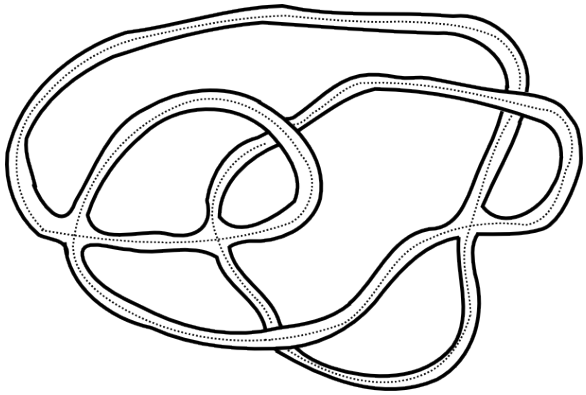


Figure 4.2: This is an orientable ribbon surface with one boundary component, which can be seen by tracing out the thickened edges of the figure. Since the Euler characteristic is -3, it must be a surface of genus 2 with one boundary component. The core curve is indicated with dotted lines.

Figure 4.2 gives an example of a filling curve with 3 self-intersections on a surface of genus 2. Now imagine cutting the ribbon surface transverse to the core curve as indicated in Figure 4.3. The hollow and filled in circles indicate how the four "loose ends" of the boundary are connected. Attaching

the additional piece given yields a surface of genus 3 and one connected boundary component. Cutting this new ribbon surface near the new added piece, we see that the four "loose ends" have the same pattern, so we may attach as many pieces as desired to yield a surface of arbitrarily higher genus with one connected boundary component. Now, we may glue a disk to this surface using the single boundary component to obtain a curve γ on a surface of genus g whose complement is a single topological disk. The last thing to check is that this curve is minimal, that is, cannot be homotoped to a curve with fewer intersections. Since the complement of γ is connected, there can be no monogons or bigons, and therefore our curve is minimal.

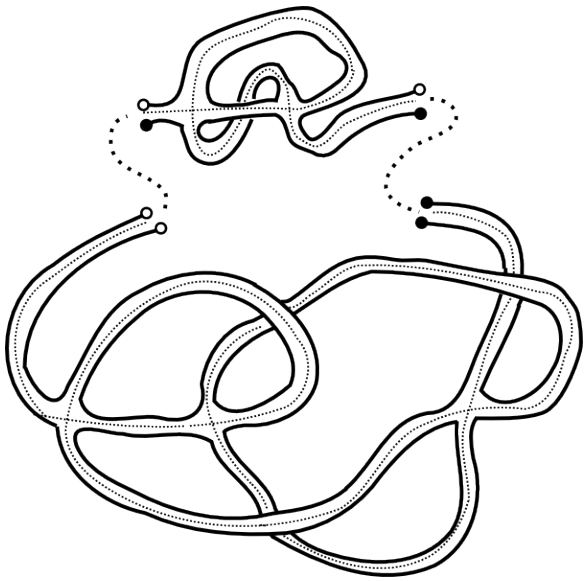


Figure 4.3: We may attach several of the pieces depicted here side by side. For each piece we attach, we increase the genus of the surface by one, and the resulting surface will still have one connected boundary component.

□

If we wish to find filling curves on a surface S_g whose complement is a pair of topological disks, we must construct curves with $2g$ self-intersections. It turns out examples exist for every genus $g \geq 2$ in this case as well, as the following argument demonstrates.

Theorem 4.3.2. *For $g \geq 2$, there exists a curve γ with $2g$ self-intersections, whose complement is a pair of topological disks.*

Proof. As in the previous theorem, the proof is inductive. Figure 4.4 demonstrates the piece to be added in each step. The only thing to prove is that these curves contain no monogons or bigons. We apply Lemmas 4.2.5 and 4.1 to the vertices of these graphs. We systematically see that all choices for vertices and paths fail one of the conditions in the lemmas. As an example, in Figure 4.4, neither of the vertices in the upper left portion of the image could be the vertex of a bigon because for each possible "sector" emanating from the vertex, there is an arc intersecting the potential bigon which fails to satisfy the first property in Lemma 4.1. □

Remark 4.3.3. Connected curves with more intersection numbers and more complementary regions may be constructed in a similar manner, but the analysis becomes more delicate and uninteresting. One performs various kinds of surgeries on the ribbon graphs and then applies Lemmas 4.2.5 and 4.1

The constructions in this chapter lead to some interesting questions for future analysis:

Question 4.3.4. Can every minimally self-intersecting filling curve on a surface S_g be obtained by surgering a minimal filling curve on S_{g-1} as in the proof of Theorem 4.3.1?

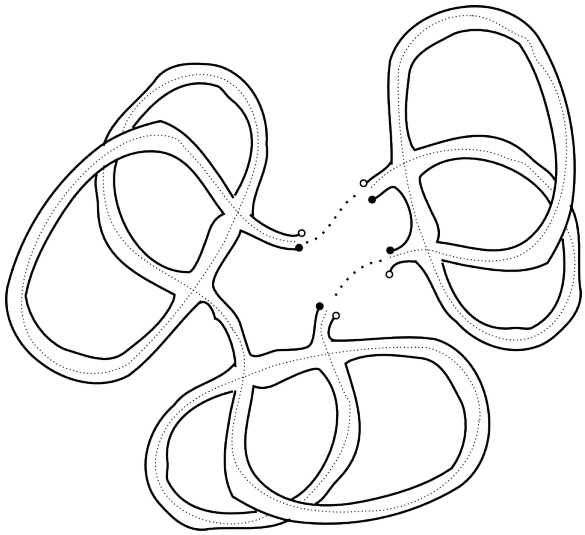


Figure 4.4: The larger piece can be glued back up to produce an example in genus 2, while additional pieces can be glued to obtain arbitrarily higher genus examples.

Question 4.3.5. How many homeomorphism classes of minimally self-intersecting filling curves are there? This number appears to grow at least exponentially, but without an answer to Question 4.3.4, we cannot obtain an upper bound.

Chapter 5

Topology informing Geometry

There are many instances in which the combinatorics of a curve (or collections of curves) can inform the geometry. A classic example of this is the so-called *collar lemma*:

Theorem 5.0.6. *Suppose α is a simple closed curve on a hyperbolic surface of length l and β is a curve that intersects α k times. Then the length of β is at least:*

$$2 \cdot arcsinh\left(\frac{1}{sinh(\frac{x}{2})}\right)$$

Corollary 5.0.7. *If α and β are two simple closed curves on a hyperbolic surface that intersect, then one of them has length at least $2 \cdot arcsinh(1)$.*

This collar theorem found in [25] was perhaps the inspiration for Thurston's famous compactification of Teichm$\acute{}$uller space using projective measured laminations ([36]). As you approach the boundary of Teichm$\acute{}$uller space, by Mumford's compactness theorem ([17]), there must be a simple closed curve

whose length is becoming arbitrarily small. By the collar lemma, any curve that intersects this curve must get arbitrarily large. If we projectivize, then the length of any curve in the limit is essentially controlled by how many times it intersects this simple closed curve which is shrinking.

The previous results mentioned produce information about *lengths* from *topological* data, namely intersection numbers. In this chapter, we will explore how the larger-scale combinatorial structure of a curve can be used to obtain information about lengths and angles of intersection.

5.1 A natural map

Let $\mathcal{P}(n)$ denote the configuration space of hyperbolic n-sided polygons with a fixed (arbitrary) labeling of the edges. This space has a natural topology - two polygons are close if the edges with the same label are close in length, and if the angles formed by edges with the same labels are close. $\mathcal{P}(n)$ can be given the structure of a $2n-3$ dimensional manifold embedded in ${\mathbb{R}_+}^{2n}$. (see [32]).

A collection of curves γ on a surface in general position is said to contain a *triangular region* if there is an embedded triangle in the surface whose sides correspond to arcs of curves in γ and whose vertices correspond to distinct points of intersection. Curves which contain triangular regions do not have a unique topological configuration: one can "slide" sides of the triangle past vertices to obtain topologically distinct configurations.

Now let γ be a collection of filling curves on $\S_g$ with k total intersections and no triangular regions. Then $\S_g \,/\, \gamma$ is a collection of hyperbolic polygons $P_0, P_1, ..., P_m$ with various numbers of sides. We consider the polygons to be *marked* which means we label each edge. Since our curves contain no

triangular regions, the topological configuration of γ is fixed, which means that the complementary regions will always have the same number of sides, independent of whatever metric is chosen on our surface.

Proposition 5.1.1. *Let s_i be the number of sides of polygon P_i in the complement of γ, which has k total intersections. Then $\sum_{i=0}^{m} s_i = 4k$.*

Proof. Since γ is filling, it determines a decomposition of $\S_g$ into k vertices, $2k$ edges, and m faces. Since each edge of this decomposition corresponds to two glued edges of the polygons we obtain the result. □

Since γ has no triangular regions, for any choice of hyperbolic structure, the complement of the geodesics maintain their number of edges, so the numbers $s_0, s_1, ...s_m$ are actually invariants of the curve γ on $\S_g$, and not invariants of the hyperbolic metric chosen on $\S_g$.

This gives us a natural map $\Phi_\gamma : Teich(\S_g) \to \prod_{i=0}^{m} \mathcal{P}_\rangle$ which sends a marked hyperbolic metric to the collection of hyperbolic polygons in the complement of γ. (We still must choose first an arbitrary labeling for the edges of the P_i, but after this choice is made, the combinatorics of γ determines this map uniquely). The collection of curves γ determines how these various polygons should be glued to one another, and which corners must be joined to form intersections.

Proposition 5.1.2. *The image of Teichmüller space under Φ_γ is the smooth submanifold of $\prod_{i=0}^{m} \mathcal{P}_\rangle$ where the identified edges (given by γ) have the same length, and where the four angles forming each vertex (again, determined by γ) alternate and add up to 2π. The dimension of this submanifold is $6g-6$.*

Proof. The description of the image submanifold is apparent, since the polygons are determined by the arcs of a closed geodesic. We will give a heuristic

argument to calculate the dimension, for brevity. By Corollary 4.1.2, we know that $m = k-2g+2$, and by Proposition 5.1.1, the total sum of sides is $4k$. Imposing a condition on a side or an angle of a polygon reduces the dimension of the configuration space by one. Since the angles of the polygons P_i are determined by the intersection angles of a geodesic, each intersection imposes a total of 3 conditions on the configuration spaces, since choosing any one of the 4 angles at an intersection determines the others. Thus we have $3k$ conditions coming from restrictions on the angles. Similarly, since edges of polygons must be glued in pairs, this imposes a further $2k$ conditions on the edge lengths. The product of configuration spaces of the P_is therefore has dimension $2(4k) - 3(k - 2g + 2) = 5k + 6g - 6$. There are $5k$ independent gluing conditions, which produces a manifold of dimension $6g - 6$. □

In fact, Φ_γ is a homeomorphism onto its image:

Proposition 5.1.3. *Φ_γ is 1-1.*

Proof. Suppose we are given a collection of labeled polygons in the image of Φ_γ. We can place these polygons, one at a time, in $\mathbb{H}^2$, using the curve γ to determine which labeled sides of the polygons must be identified with one another. This collection of glued polygons determines a fundamental domain and thus a hyperbolic structure. □

5.2 Applications of Φ_γ

In this section, we will showcase two applications of the function Φ_γ which can be used to uncover information about curves. First, we will examine the filling curve produced in Theorem 4.3.1 and compute a sharp lower bound

on its length. The main observation is that if a filling curve has the minimal number of self-intersections, then the complement is a single polygon. The map Φ_γ is quite simple in this case. Next, will look at Φ_γ for punctured surfaces, and use that map to make a statement about angles of intersections between curves.

5.2.1 A lower bound on length

We begin with the following fact:

Proposition 5.2.1. *A regular k-sided hyperbolic polygon with area A has perimeter given by*

$$P(A,k) = k \cdot cosh^{-1}\left(\frac{cos\left[\frac{(k-2)\pi - A}{2k}\right]^2 + cos\left[\frac{2\pi}{k}\right]}{sin\left[\frac{(k-2)\pi - A}{2k}\right]^2}\right)$$

Proof. This is a consequence of the special hyperbolic law of cosines. A regular hyperbolic polygon of area A with k sides can be partitioned into k isosceles triangles whose angles are $\frac{2\pi}{k}$ and $\frac{(k-2)\pi - A}{2k}$. Summing up the length of the edges contributing to the perimeter of the regular polygon, we obtain the result. □

A minimally intersecting filling curve on a genus g (g $\geq$ 2) surface produces a hyperbolic polygon of area $-\chi(S_g) = 2\pi(2g-2)$ with $8g-4$ sides. This leads to the following:

Theorem 5.2.2. *A minimally intersecting filling curve on a surface S_g must have length at least half that of a regular right-angled $(8g-4)$-gon. This lower bound is:*

$$(4g-2)\cdot cosh^{-1}\left(2\cdot cos\left[\frac{\pi}{4g-2}\right]+1\right)$$

Proof. Among all polygons with k sides, enclosing a fixed area A, the regular polygon minimizes the perimeter. For a proof, see [9]. So suppose we have a filling curve γ, which has $2g-1$ intersections. Then the complement of the curve γ is a hyperbolic polygon, with a total of $8g-4$ sides. The regular polygon with this area and number of sides is right-angled, which can be seen by the angle-defect formula. The length of γ is half this total perimeter, since edges of the polygons are identified in pairs. Plugging in these parameters to the previous proposition, we obtain the result. □

Asymptotically, this length grows linearly in g, and thus linearly in the number of intersections. Contrast this with the results in [8], which say that the shortest length of a typical curve with k intersections grows like $\sqrt{k}$.

5.2.2 Angles of intersection

The map Φ_γ has a natural definition for geodesics on *punctured surfaces*. One just has to allow the various polygons in the complement to be punctured. These polygons also have a natural configuration space, so the definition of Φ_γ holds in this setting as well.

For small numbers of sides, direct calculations about angles and sides of polygons can be made. One such example is the following theorem from [6], which we will make use of:

Theorem 5.2.3. *Let $\mathcal{Q}$ be a hyperbolic quadrilateral with one angle ϕ and three right angles. As arranged in Figure 5.1, the following relationships hold:*

- $sinh(x)sinh(y) = cos(\theta)$
- $cosh(x) = cosh(z)sin(\theta)$.

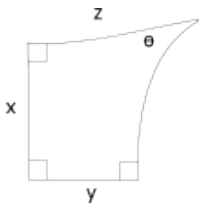


Figure 5.1: The quadrilateral for Theorem 5.2.3.

Theorem 5.2.3 allows us to explicitly determine the space $\mathcal{C}(P_4)$ (as a subset of Euclidean space) in a variety of situations.

In [30] and [31], it is shown that angles of intersection between certain collections of geodesics can in some cases be used to parameterize the global hyperbolic metric on a surface. The results involve various computations using traces and properties of $SL(2, \mathbb{R})$. Here, we adopt a more constructive approach to show that certain collections of curves *fail* to parameterize the global hyperbolic metric on a surface. The methods developed here can be used to recover some of the results in [30] and [31] and highlight some interesting complexities in studying angles of intersection.

Let α and β be two simple closed curves on the punctured torus which intersect twice. We will study how the two angles of intersection change as we deform the hyperbolic metric on the surface. One might naively expect these two angles of intersection to completely determine the hyperbolic metric, but the reality is a little subtler. First we begin with a useful proposition which will allow us to "normalize" many arguments that will follow.

Proposition 5.2.4. *Let α and β be two simple closed curves on the punctured torus which intersect twice, and let α' and β' be another such pair of*

curves. Then there is a homeomorphism of the surface ϕ which sends $\alpha \cup \beta$ to $\alpha' \cup \beta'$.

Proof. We will show that the complements of any pairs of curves satisfying the hypothesis are topologically identical, which will allow us to construct a homeomorphism from one pair of curves to the other. First, cut the two punctured tori T and T' along α and α'. Since the curves are simple, the result is a topological pair of pants in both cases. Now, we know that β and β' intersect the paired curve exactly twice, for each pair of pants, there are two arcs connecting the two boundary components coming from the first cut. Pick one of these arcs on both surfaces and cut along them. The result is a punctured disk in both cases, which is shown in Figure 5.2. The last arc has two topologically distinct ways of traversing this punctured disk (the solid and dashed lines in the rightmost figure). However, they differ by the homeomorphism which rotates the disk. Using Proposition 2.3.1, one can work backwards from the punctured disk and use the various homeomorphisms between the intermediate surfaces to give a homeomorphism of the punctured torus which sends the curves $\alpha \cup \beta$ to $\alpha' \cup \beta'$. □

From the above proof, one sees that any pair of geodesic curves on a hyperbolic punctured torus satisfying the hypothesis of the proposition partitions the punctured torus into two components: a quadrilateral with alternating angles, and a punctured quadrilateral with alternating angles.

Proposition 5.2.5. *A quadrilateral has alternating angles if and only if it has alternating side lengths.*

Proof. Let x_1, x_2, x_3, x_4 be lengths of the 4 sides of the quadrilateral. The space of quadrilaterals with angles $\theta_1, \theta_2, \theta_1, \theta_2$ arranged cyclically, is a 1-dimensional submanifold of $\mathbb{R}^4$. The permutation of $\mathbb{R}^4$ switching the sides

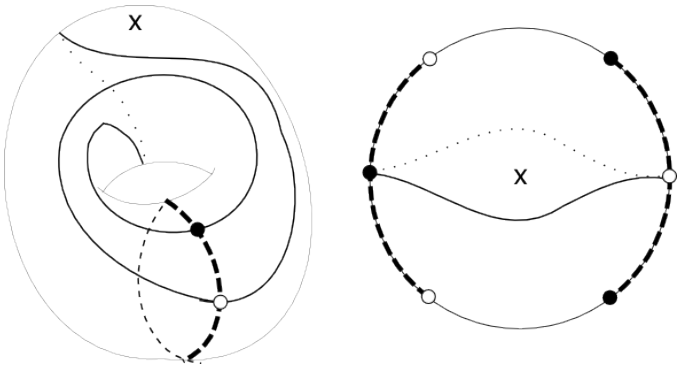


Figure 5.2: A pair of curves as in Proposition 5.2.4, along with the surgered surface described in the proof of the theorem.

x_1 with x_3, and x_2 with x_4 preserves this manifold. This tells us that if we graph x_3 as a function of x_1 along this submanifold, the graph is symmetric about the line $x_1 = x_3$. If $x_1 \neq x_3$ identically, then as x_1 approaches ∞, we must have x_3 approaching 0. This is clearly impossible, since x_3 increases with x_1. The same argument works for x_2 and x_4. Now for the converse. Divide the quadrilateral into two triangles using one of the diagonals. These triangles are isometric since they have the same side lengths. Thus the two angles on either side of the diagonal are equal. Choosing the other diagonal, we see that the other pair of angles are likewise equal. □

Theorem 5.2.6. *The space of alternating quadrilaterals is homeomorphic to an open subset of* $\mathbb{R}^3$ *given by sending a quadrilateral* $\mathcal{Q}$ *to the triple* (x, θ_1, θ_2)*, where* x *is the length of a side, and* θ_1 *and* θ_2 *are the alternating angles.*

Proof. We will show that each triple in the image determines a *unique* quadrilateral, which will prove the theorem. Consider a hyperbolic geodesic A in the disk model passing through the original. Construct another geodesic

B through the origin forming an angle of θ_1 with the first geodesic. Go out a distance x along the first geodesic and construct a geodesic D forming an angle of θ_2. We now have constructed three of the four geodesics which form the sides of the quadrilateral. Consider a a geodesic ray C emanating from B with angle θ_2, whose basepoint is free to slide along B. If this basepoint is very far from the origin, the ray will not intersect C. In the range for which it does intersect C, we see that the angle of intersection with D is a continuous and monotone function, with limits 0 and π. Thus, by the intermediate value theorem, there is a unique basepoint for which the angle is θ_1, which completes the quadrilateral. □

Since the Teichmüller space of the punctured torus is 2-dimensional, and the space of alternating quadrilaterals is 3-dimensional by Theorem 5.2.6, we may ask which alternating quadrilaterals appear in the image of Φ.

Definition 5.2.7. A quadrilateral whose sides are labeled $a_0, .., a_3$ with $length(a_0) = length(a_2) = x$ and $length(a_1) = length(a_3) = y$ is called a *good quadrilateral with respect to side* a_0 if the orthosegment between the two geodesics containing a_0 and a_2 is $arcsinh(\frac{1}{sinh(x)})$, the collar function evaluated at x. See Figure 5.3

Proposition 5.2.8. *The space of good quadrilaterals is homeomorphic to* $\mathbb{R}_+ \times \mathbb{R}$.

Proof. From a pair $(x, y) \in \mathbb{R}_+ \times \mathbb{R}$, we construct a unique good quadrilateral as follows: Choose an oriented geodesic $\mathcal{X}$ in $\mathbb{H}^2$ and pick a hyperbolic line segment of length x, whose endpoints are labeled x_0 and x_1. Take an orthosegment $\mathcal{O}$ of length $arcsinh(\frac{1}{sinh(x)})$, whose basepoint is at a distance y from x_1, on the "left" side of the oriented geodesic containing x_0 and x_1.

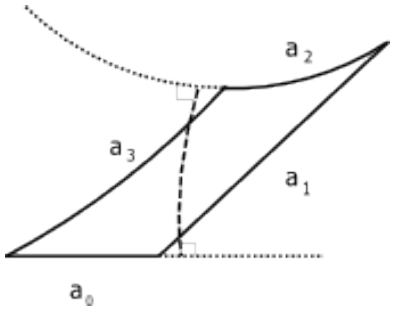


Figure 5.3: A quadrilateral is *good* if the length of the perpendicular connecting opposite sides is the collar length of the sides it connects.

There is another oriented geodesic $\mathcal{Y}$, which is uniquely determined by our placement of $\mathcal{O}$ such that $\mathcal{O}$ is "on the left" with respect to the orientation. Now choose two points y_0 and y_1 on $\mathcal{Y}$ such that the signed distances between x_1 and $x_0$with the basepoint of $\mathcal{O}$ on $\mathcal{X}$ are the same as the signed distances between y_0 and y_1 on the basepoint of $\mathcal{O}$ on $\mathcal{Y}$. This condition uniquely determines y_0 and y_1. The four points x_0, x_1, y_0, y_1 then determine a good quadrilateral, by construction. This process can clearly be reversed, yielding a unique pair of coordinates (x, y) from any good quadrilateral with labeled sides. □

Definition 5.2.9. A good quadrilateral with respect to side a_i is said to have *no twist* if the orthosegment connecting side a_i with side a_{i+2} (mod 4) passes through the diagonal vertices.

Theorem 5.2.10. *If $\mathcal{Q}$ is good with respect to some side a_0, then $\mathcal{Q}$ determines a unique hyperbolic structure for the punctured torus.*

Proof. Consider the good quadrilateral $\mathcal{Q}'$ obtained by sliding the segment a_2 along the geodesic it lies on, until $\mathcal{Q}'$ has no twist. We will first show that $\mathcal{Q}'$ determines a hyperbolic structure for a punctured torus. Let A'

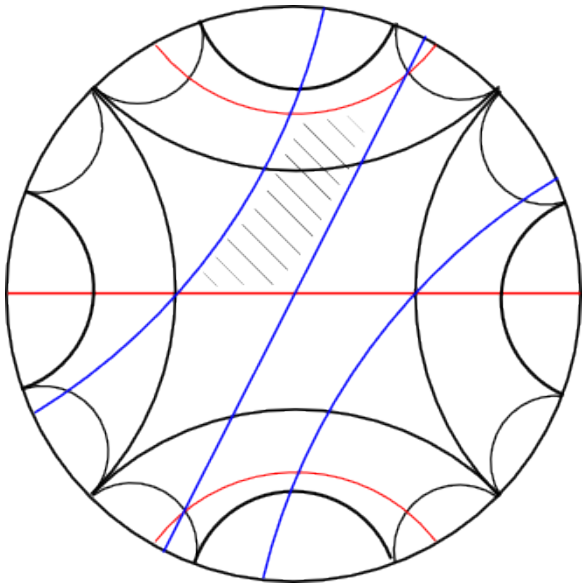


Figure 5.4: Some lifts of the curves α and β to the universal cover, and the alternating quadrilateral they determine.

be the loxodromic isometry which sends one vertex of a_0 to the other (note that this isometry is the same for $\mathcal{Q}$ and $\mathcal{Q}'$. Let B be the loxodromic isometry which sends the basepoint of the orthosegment between a_0 and a_2 of $\mathcal{Q}'$ to the "top" of the orthosegment. Note that the axes of A and B are orthogonal, since $\mathcal{Q}'$ has no twist.

We will now construct a fundamental domain for the action of $G = \langle A = A'^2, B \rangle$. Consider the orthogeodesic emanating from the corner of $\mathcal{Q}'$ which is not the axis of B. Apply A' twice to this orthogeodesic to obtain another geodesic orthogonal to the axis of A. The region in between these two orthogeodesics is a fundamental region for the action of A on $\mathbb{H}^2$. Consider the geodesics that are orthogonal to the axis of B and precisely $\frac{1}{2}$ the translation length of B, in both directions from the axis of A. The region between these two orthogesics is a fundamental region for the action

of B on $\mathbb{H}^2$. The intersection of these two fundamental regions is an ideal hyperbolic quadrilateral, since we have constructed the side lengths such that they satisfy the condition in Theorem 5.2.3 when $\theta = 0$. The isometries A^2 and B identify opposite sides of this ideal quadrilateral, so that the final object is a punctured torus endowed with a hyperbolic structure. $\mathcal{Q}$ defines a hyperbolic structure by performing a Nielsen twist about the axis of A until $\mathcal{Q}'$ is deformed back into $\mathcal{Q}$. □

We can actually use Theorem 5.2.10 to prove the following symmetry result about good quadrilaterals, and provides another interesting way in which topology determines geometry:

Proposition 5.2.11. *If $\mathcal{Q}$ is a good quadrilateral with respect to side a_i, then it is also good with respect to side a_{i+1}.*

Proof. Since $\mathcal{Q}$ is good, it determines a hyperbolic structure generated by two loxodromic transformations A and B. Note that the axes of these transformations project to simple closed geodesics on the punctured torus. Using the definitions of A and B in the proof of Theorem 5.2.10, we see that the sides of $\mathcal{Q}$ are coming from lifts of the geodesics A, and B^2A. Since B^2A is another simple closed geodesic, we know that there is a homeomorphism ϕ of the punctured torus sending the curve A to B^2A. We can further choose ϕ so that this homeomorphism sends B to BA. This has the effect of switching the roles of of the even and odd indexed sides of the quadrilateral $\mathcal{Q}$. Since all we've done is change the marking, the hyperbolic structure remains intact, and this forces the orthosegment between the appropriate sides to have the proper length. □

Theorem 5.2.10 tells us that the space of good quadrilaterals $\mathcal{GQ}$ (with

the obvious topology) parameterizes Teichmüller space. Consider the following two operations f and g on $\mathcal{GQ}$, based on Figure 5.5:

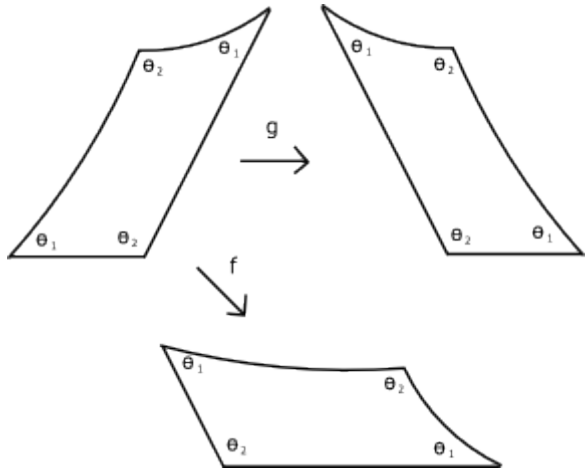


Figure 5.5: The natural geometric operations on a quadrilateral Q.

- f cyclically permutes the labels of the sides, in effect rotating the quadrilateral.
- g switches the locations of the two angles.

These two operations can be described in terms of the coordinate system we have placed on $\mathcal{GQ}$. If $\mathcal{Q}$ is given by a point in $\mathbb{R}_+ \times \mathbb{R}$ as in Proposition 5.2.8 then $g(x, y) = (x, -(x + y))$. In terms of these coordinates, the expression for f is more complicated, depending on various compositions of hyperbolic and ordinary trigonometric functions.

Proposition 5.2.12. *$g \circ f$ has a unique fixed point.*

Proof. The composition has the effect of reflecting the quadrilateral, switching the roles of the upper left point of the quadrilateral and the lower right.

The quadrilateral is fixed by the action if and only if the side lengths are the same and the angles are the same. There is a 1- parameter family of such quadrilaterals, but notice that only one of them satisfies the collar equation. □

Theorem 5.2.13. *Let $\mathcal{Q}$ be a good quadrilateral. Then $g \circ f(\mathcal{Q})$ is a good quadrilateral with the same angles.*

Proof. By Theorem 5.2.11, $f(\mathcal{Q})$ is still a good quadrilateral, and since g only changes the twist parameter, $g \circ f(\mathcal{Q})$ is also a good quadrilateral. f switches the orderings of the sides of $\mathcal{Q}$ and sends $\mathcal{Q}$ which has angles (θ_1, θ_2) to a good quadrilateral with angles (θ_2, θ_1). Then g switches the roles of the angles again, producing another good quadrilateral whose alternating angles match the angles of $\mathcal{Q}$. □

The above theorem tells us that we can't use angles of intersection alone (at least for the curves α and β that we've chosen) as global parameters for Teichmüller space. However, the next theorem tells us that the set of points in Teichmüller space where the angles of intersection fail to give local parameters is a discrete set.

Theorem 5.2.14. *Let α and β be any two simple curves which intersect minimally twice and fill on the punctured torus. Then the two angles of intersection give local parameters for Teichmüller space, except at a discrete set of points.*

Proof. Let us recap what we have done so far, which will guide the proof. We can embed the space of alternating hyperbolic quadrilaterals in $\mathbb{R}^3$ using coordinates (x, θ_1, θ_2), corresponding to an alternating quadrilateral with alternating angles θ_1 and θ_2 with two side lengths x. Teichmüller space

can then be embedded as the triples (x, θ_1, θ_2) for which the perpendicular connecting the two sides of length x is the collar function of x. The perpendicular length $P(x, \theta_1, \theta_2)$ is an analytic function of x, θ_1, and θ_2, being composed of various trigonometric and hyperbolic functions. If we fix a pair of angles θ_1 and θ_2, we obtain an an analytic function $p(x)_{\theta_1,\theta_2}$ of a single variable. The collar function is also an analytic function of x.

Suppose for contradiction that for every neighborhood U_i of some point x in the embedded image, there is an $x_i \in U_i$ with the same angles. The identity theorem for analytic functions says that if two analytic functions are equal on a set of points that accumulate in the domain, then they are actually equal globally throughout the domain. But $P(x)$ and the collar function are not globally equal. This shows that there is a neighborhood of x for which x is the only point that has θ_1, θ_2 as the angles. This is almost what we need, but not quite; We must prove that the points where the functions and their derivatives are equal is discrete, for if the functions are equal but their derivatives are not, we know there is a neighborhood where *every* pair of angles is distinct, via the implicit function theorem. This result can be obtained simply by implying the same identity theorem to the derivatives. □

It is conjectured that the fixed point of $g \circ f(\mathcal{Q})$ is the *only* point where the angles of intersection fail to give local parameters, but that requires a more careful analysis to fully justify.

5.3 Future Questions

There are various intertwined geometric and topological notions attached to a free homotopy class. Absent a hyperbolic metric, one can discuss intersections and configurations, for example. In the presence of a hyperbolic metric, one can also ask questions about length and angle of intersection as well. We may assign values to one or more of these various properties of curves and ask which subsets of Teichmüller space preserve these values? Sometimes the answer is not interesting. For example, every geodesic realizes the minimal number of intersections. Sometimes, the answer is more subtle: If a collection of curves has triangular regions, then the curve has two or more topologically distinct configurations. In [22], it is shown that for certain curves, certain topological configurations are never realized by a geodesic on a hyperbolic surface. For a configuration that is realizable, what is the topology of the subset of Teichmüller space which preserves the configuration?

If a curve γ on S_g is filling, then a result in [26] states that there is a unique point in Teichmüller space which minimizes the length of γ. If we choose any L larger than this minimal length, then the subset of Teichmüller space for which the length of γ is L is homeomorphic to S^{6g-7}. This is due to the fact that the length of a filling curve is convex and limits to infinity along an earthquake path, and earthquake paths are parameterized by $\mathcal{PML}$, the space of projective measured laminations, which is homeomorphic to S^{6g-7}.

What about the subset of Teichmüller space which preserves a tuple of intersection angles for a filling curve? We have just analyzed a special case of this on the punctured torus. The answer for a general curve on $S_{g,n}$ will surely be more complicated. The following heuristic observation suggests

that this may be a very interesting set in some cases.

Observation 5.3.1. Let γ_g be a minimally self-intersecting filling curve on S_g, as in Theorem 4.3.1, and let $\Theta = (\theta_1, ..., \theta_{2g-1})$ be a tuple of intersection angles realized in some hyperbolic metric. Then the set of Teichmüller space which preserves Θ should have dimension $4g - 5$.

Proof. (heuristic) A filling geodesic with $2g-1$ self-intersections has an $8g-4$ sided hyperbolic polygon as its complement. If we want to fix the angles of intersection for our geodesic, then the angles of this $8g - 4$-gon are also fixed. The space of hyperbolic $4g-4$-gons with fixed angles is a manifold of dimension $8g-7$. Since our polygon is coming from a geodesic, the lengths of the sides of the complementary polygon must match in pairs, which reduces the dimension by $4g - 2$. Thus, assuming various transversality conditions, the space of polygons producing a geodesic with intersection angles Θ should be $4g - 5$ dimensional. □

The number $4g - 5$ comes up in the literature as the conjectural lower bound for the dimension of a deformation retract of moduli space (see [19],[24]). Perhaps this is just a coincidence, but it seems worth a more detailed analysis in the future.

Bibliography

[1] W. Abikoff, *"The Real analytic theory of Teichmüuller space."* Lecture notes in Math.; 820 (1989).

[2] J. W. Anderson, *Hyperbolic geometry*. Second edition. Springer Undergraduate Mathematics Series. Springer-Verlag London, Ltd., London, 2005. xii+276 pp.

[3] K. Ahara and Y. Yamazaki. *"Software and Algorithm of Drawing Surfaces."* Interdisciplinary Information Sciences 9.1 (2003): 157-174.

[4] T. Aougab and S. Huang. *"Minimally intersecting filling pairs on surfaces."* Algebraic and Geometric Topology 15.2 (2015): 903-932.

[5] L. Ahlfors and C. Earle. *"Lectures on quasiconformal mappings."* (1966): 3.

[6] A. F. Beardon, *The geometry of discrete groups*, Graduate Texts in Mathematics, 91. Springer-Verlag, New York, 1983. xii+377 pp.

[7] A. Basmajian, *The stable neighborhood theorem and lengths of closed geodesics.* Proceedings of the American Mathematical Society 119.1 (1993): 217-224.

[8] A. Basmajian, Ara. *"Universal length bounds for non-simple closed geodesics on hyperbolic surfaces."* Journal of Topology (2013): jtt005.

[9] K. Bezdek, K. *"Ein elementarer Beweis fr die isoperimetrische Ungleichung in der Euklidischen und hyperbolischen Ebene."* Ann. Univ. Sci. Budapest. Etvs Sect. Math 27 (1984): 107-112.

[10] J. Birman and C. Series, *An algorithm for simple curves on surfaces*, J. London Math. Soc. Vol 29 Iss 2 (1984), 331-342.

[11] P. Buser, *Geometry and spectra of compact Riemann surfaces*. Progress in Mathematics, 106. Birkhuser Boston, Inc., Boston, MA, 1992. xiv+454 pp.

[12] M. Chas, *Combinatorial Lie Bialgebras of curves on surfaces*, Topology Vol 43 Iss 3 (2004), 543-568

[13] M. Chas, *Minimal intersection of curves on surfaces*, Geometriae Dedicata Vol 144 (2010), 25-60

[14] M. Chas, Moira and A. Phillips. *"Self-intersection numbers of curves on the punctured torus."* Experimental Mathematics 19.2 (2010): 129-148.

[15] D. Chillingworth, *Simple closed curves on surfaces*, Bull. London Math. Soc. Vol 1 (1969), 310-314

[16] M. Cohen and M. Lustig, *Paths of geodesics and geometric intersection numbers I*, Combinatorial Group Theory and Topology, Ann. of Math. Studies Vol 111 (1987), 479-500.

[17] B. Farb and D. Margalit, *A Primer on Mapping Class Groups* (PMS-49). Princeton University Press, NJ, 2011.

[18] H. Farkas and I. Kra. *Riemann surfaces*. Springer New York, 1992.

[19] J. Harer, *The virtual cohomological dimension of the mapping class group of an orientable surface* Inventiones mathematicae 84.1 (1986): 157-176.

[20] J. Hass, P. Scott, *Intersections of curves on surfaces*, Israel J. Math. Vol 51 (1985), 90-120

[21] J. Hass, P. Scott, *Shortening curves on surfaces*, Topology Vol 33 No 1 (1994), 25-43

[22] J. Hass, P. Scott, *Configurations of curves and geodesics on surfaces*, Geometry and Topology Monographs 2 (1999): 201-213.

[23] A. Hatcher, *Algebraic topology*. 2002. Cambridge UP, Cambridge 606.9.

[24] L. Ji and S. Wolpert. "*A cofinite universal space for proper actions for mapping class groups*." the tradition of Ahlfors-Bers. V (2008): 151-163.

[25] L. Keen *"Collars on Riemann surfaces."* Discontinuous groups and Riemann surfaces (Proc. Conf., Univ. Maryland, College Park, Md., 1973). 1974.

[26] S. Kerckhoff, *"The Nielsen realization problem."* Annals of mathematics (1983): 235-265.

[27] G. Martin, *The persistence of angle*. Centre for Mathematics and its Applications, School of MAthematical Sciences, Australian National University, 1993.

[28] B. Maskit, *Kleinian groups*, Springer-Verlag, Berlin, 1988. xiv+326 pp.

[29] J. Munkres, *Elements of algebraic topology.* Vol. 2. Reading: Addison-Wesley, 1984.

[30] Y. Okumura, *"On the global real analytic coordinates for Teichmller spaces."* Journal of the Mathematical Society of Japan 42.1 (1990): 91-101.

[31] Y. Okumura, *"Global real analytic length parameters for Teichmller spaces."* Hiroshima Mathematical Journal 26.1 (1996): 165-179.

[32] J. Porti, *"Hyperbolic polygons of minimal perimeter with given angles." Geometriae Dedicata 156.1 (2012): 165-170.*

[33] B. L. Reinhart, Algorithms for Jordan curves on compact surfaces, Annals of Math. Vol 75 (1962), 209-222

[34] J. Stillwell, *Classical topology and combinatorial group theory.* Vol. 72. Springer Science and Business Media, 2012.

[35] J. Stillwell, *Geometry of surfaces.* Springer Science and Business Media, 2012.

[36] W.P. Thurston, *"On the geometry and dynamics of diffeomorphisms of surfaces."* Bulletin (new series) of the american mathematical society 19.2 (1988): 417-431.

[37] W.P. Thurston, *The geometry and topology of three-manifolds.* Princeton: Princeton University, 1979.

[38] W.Thurston, *Minimal stretch maps between hyperbolic surfaces*, arXiv.org; math; arXiv:math/9801039v1, (1998), 1-53.