GEODESICS WITH TWO SELF–INTERSECTIONS ON THE PUNCTURED TORUS

SUSAN DZIADOSZ, THOMAS INSEL, PETER WILES

1. INTRODUCTION

We will be investigating closed loops with two self-intersections on the punctured torus. Our goal in this paper is to classify these loops, and then to determine which correspond to classes of hyperbolic geodesics. Some of the techniques we will use are adapted from those laid out by David Crisp in his Ph.D. thesis. The following survey of background material includes an abbreviated version of some relevant material from the introduction to [C].

In order to examine the punctured torus, we must review some hyperbolic geometry. The upper half plane model of the hyperbolic plane, $\mathbf{H}$, is defined on the set $\{x+iy : y > 0\}$. In the hyperbolic plane, geodesics are represented by semi-circles centered on the real axis and by infinite vertical lines. Note that the choice of a pair of points on the real axis uniquely determines a geodesic on $\mathbf{H}$ up to orientation. We call these points the feet of a geodesic.

We will be using the group

$$SL(2,\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,b,c,d \in \mathbb{Z},\ ad-bc=1 \right\}$$

to act upon $\mathbf{H}$ through the homomorphism defined by

$$T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad \longmapsto \qquad Tz = \frac{az+b}{cz+d}.$$

This group of fractional linear transformations is $\Gamma = PSL(2,\mathbb{Z})$. We will denote both a matrix in $SL(2,\mathbb{Z})$ and a transformation in $PSL(2,\mathbb{Z})$ by the same symbol. Let Γ' be the commutator subgroup of Γ. Γ' is a free group on the two generators

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

We lose no generality looking at the matrix representations because Γ' is isomorphic to the commutator subgroup of $SL(2,\mathbb{Z})$.

Research conducted during the Summer 1994 NSF REU Program at Oregon State University. Thanks to Professor Tom Schmidt and Professor Dennis Garity for telling us about cool Oregonian festivals and advising us on this project. Figures 1.1, 2.1, 3.7 and 3.9 were generated with software written at the Geometry Center, University of Minnesota.

The axis of a word $W \in \Gamma'$ is the unique geodesic $\tilde{\gamma}$ in $\mathbf{H}$ which is fixed by W. W translates $\tilde{\gamma}$ along itself by a fixed hyperbolic distance. We can partition $\tilde{\gamma}$ into a fundamental segment of the form $[z_0, Wz_0]$, the directed segment from z_0 to Wz_0. Note the axes of A and B are drawn on Figure 1.1. We will call the point of intersection of these two axes i.

Elements in Γ can be either elliptic, parabolic or hyperbolic. These can be distinguished by their trace. Let $W = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. If $|a+d| = 2$, W is parabolic and has only one real solution, but the axis of W is not a vertical line. Therefore, the axis of W cannot be a geodesic. We can show that Γ' has only one class of parabolic elements, the conjugacy class of $(ABA^{-1}B^{-1})^n$. If $|a + d| > 2$, W is hyperbolic and has two real solutions. Thus, the axis of W is a geodesic in $\mathbf{H}$. Elliptic elements have finite order, but the subgroup Γ' has no elements of finite order. Thus, it has no elliptic elements and the case $|a + d| < 2$ does not occur.

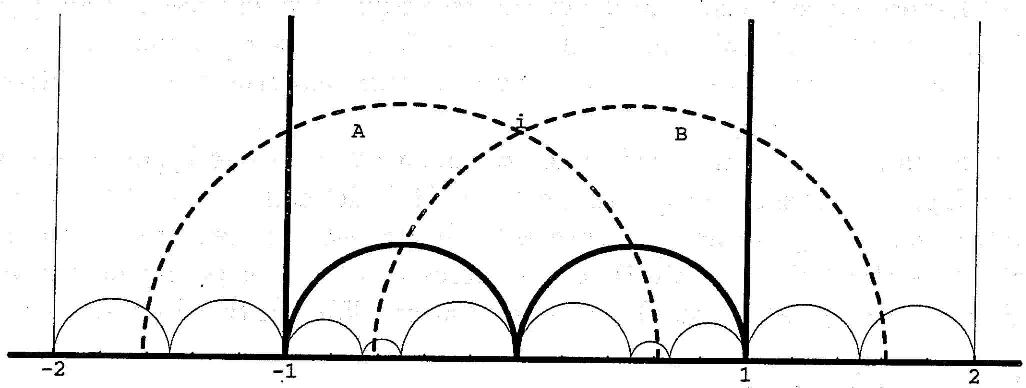

FIGURE 1.1. Copies of fundamental region $\mathcal{D}$ with the axes of A and B intersecting at the point i.

Taking the quotient group $\mathbf{H}/\Gamma'$, we obtain a group of symmetries on the hyperbolic plane. A fundamental region for the group $\mathbf{H}/\Gamma'$ is the set $\mathcal{D}$ such that each transformation g in the group Γ' carries $\mathcal{D}$ into a disjoint copy $g\mathcal{D}$ and likewise each point in the plane is carried to some point in $\mathcal{D}$ by some element in Γ' [Se2]. See Figure 1.1 for an illustration of $\mathcal{D}$. On $\mathbf{H}$, there exist an infinite number of copies of this fundamental region. With the operators A and B we can tesselate the entire upper half plane with images of $\mathcal{D}$. The operations A and B act on $\mathcal{D}$ by identifying opposite edges of the region. By making this identification, we can proceed to construct a torus in the normal fashion. Note, though, that each corner of a copy of $\mathcal{D}$ and every rational real point in $\mathbf{H}$ is identified with infinity. All such points are identified together in $\mathbf{H}/\Gamma'$, and thus the torus will contain a point at infinity, or a "puncture." We will denote the punctured torus $\mathbf{H}/\Gamma'$ as $\mathbf{T}$. The tesselation of $\mathcal{D}$ forms a covering space of $\mathbf{T}$. We can now define a projection map

$$\sigma\colon\ \mathbf{H} \longrightarrow \mathbf{H}/\Gamma' = \mathbf{T},$$

and this map is a universal covering of $\mathbf{T}$.

We can project a geodesic $\tilde{\gamma}$ on $\mathbf{H}$ onto a path γ on $\mathbf{T}$. Note that γ is covered by any fundamental segment of $\tilde{\gamma}$. We call γ together with the parameterization it inherits from any fundamental segment of $\tilde{\gamma}$ with respect to W a closed geodesic on $\mathbf{T}$ defined by the conjugacy class $[W] \in \Gamma'$.

In order to study the closed paths on $\mathbf{T}$, we make use of the algebraic structure of the path homotopy classes in $\mathbf{T}$. For a general space X and a point $x_0 \in X$, a closed oriented path in X whose initial and final point is x_0 is called a loop based at x_0. The composition of loops l_1 and l_2 is defined to be the path obtained by first following the path l_1 and then l_2, denoted $l_1 l_2$. The path homotopy classes of loops in X with base point x_0 form a group under composition of loops. This group is called the fundamental group of the space X relative to the base point x_0 and is denoted $\pi_1(X, x_0)$. Because $\mathbf{T}$ is path connected, $\pi_1(\mathbf{T}, x_0)$ is isomorphic to $\pi_1(\mathbf{T}, x)$ for any other x on $\mathbf{T}$. Therefore, we will refer simply to $\pi_1(\mathbf{T})$, and will use free (no fixed point) homotopy classes for much of our work. For a more detailed introduction to the fundamental group, see [M] and [St].

Lifts are defined in [M]. For our purposes, given a curve c in $\mathbf{H}$, a lift of c is a copy of c under some word in Γ'. All such copies map to the same $\sigma(c)$ on the torus $\mathbf{T}$. The projection σ and a fixed lift in $\mathbf{H}$ of the base point $\sigma(i)$ of $\pi_1(\mathbf{T})$ determine an isomorphism between Γ' and $\pi_1(\mathbf{T})$. We denote this isomorphism by θ and label the images of A and B by a and b, the loops which generate $\pi_1(\mathbf{T})$. Thus,

$$\theta\colon \quad \Gamma' = F(A, B) \quad \longrightarrow \quad F(a, b) = \pi_1(\mathbf{T}),$$

with $\theta(A) = a$ and $\theta(B) = b$. Just as Γ' is a free group on two generators, so is $\pi_1(\mathbf{T})$.

2. Simple Loops and Single Self-Intersections

Crisp adapts Birman and Series's results about simple curves to the punctured torus $\mathbf{T}$, giving

Theorem 2.1 (Birman and Series). *The conjugacy class of a simple loop l on $\mathbf{T}$ is either*

(a) *the identity, and l bounds a disk,*
(b) *one of $[aba^{-1}b^{-1}]$ or $[bab^{-1}a^{-1}]$, and l bounds a punctured disk or*
(c) *$[w]$ where w is a generator of $\pi_1(\mathbf{T})$, and l does not separate $\mathbf{T}$.*

It is worthwhile to consider an example of one of these conjugacy classes in more detail. In particular, we would like to examine a class that contains geodesics. We know that the conjugacy classes $[aba^{-1}b^{-1}]$ and $[bab^{-1}a^{-1}]$ correspond to parabolic classes in Γ', but generators correspond to hyperbolic classes. Therefore, consider a geodesic γ in the class $[a]$ on $\mathbf{T}$ and choose a word in $\theta^{-1}([a])$. We will use A because it is the simplest such, but any conjugate will work. Note that A fixes an axis $\tilde{\gamma}$ in $\mathbf{H}$ with feet $p_1, p_2 = (-1 \pm \sqrt{5})/2$. Since A transforms a point on the left edge of $\mathcal{D}$ to a point on the bottom-right edge of $\mathcal{D}$, a fundamental segment for $\tilde{\gamma}$ is its intersection with the fundamental region $\mathcal{D}$. We will call this segment $\mathcal{S}$, and orient it from left to right. We know that γ is covered by and has the same orientation as the projection of $\mathcal{S}$ onto $\mathbf{T}$ [C].

To completely understand the path of γ on $\mathbf{T}$, it is sufficient to examine the lifts of $\mathcal{S}$ into $\mathcal{D}$. However, all of $\mathcal{S}$ is already inside $\mathcal{D}$, so there is only the identical lift to consider. We can see the axis of A in the fundamental region $\mathcal{D}$ in Figure 1.1..

In his thesis, Crisp investigated the properties of geodesics with one self-intersection and applied them towards finding isolated values in the Markoff spectrum. To classify loops with single self-intersection, he considered the composition of two simple loops with transverse intersection and demonstrated the following theorem.

Theorem 2.2 (Crisp). *The conjugacy class in $\pi_1(\mathbf{T})$ of a loop on $\mathbf{T}$ with a non-trivial single self-intersection is either*

(a) $[(aba^{-1}b^{-1})^2]$ *or* $[(bab^{-1}a^{-1})^2]$ *or*
(b) $[g(a^2)]$ *or* $[g(abab^{-1})]$ *or* $[g(aaba^{-1}b^{-1})]$ *for some* $g \in Aut\ \pi_1(\mathbf{T})$.

Conversely, each of these conjugacy classes contains such a loop.

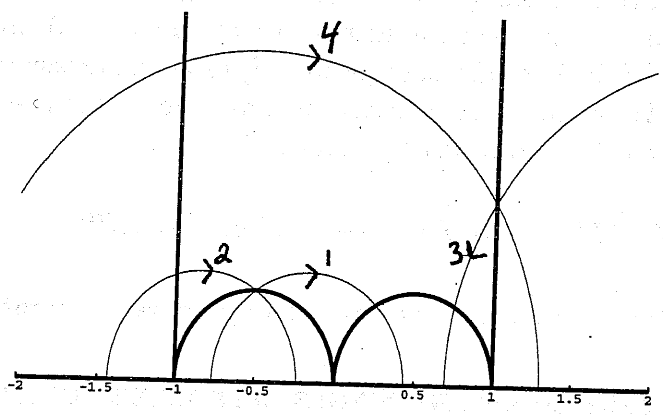

FIGURE 2.1. Lifts of the axis of $ABAB^{-1}$ in the fundamental region $\mathcal{D}$.

Now, let us consider an example of a geodesic with one self-intersection. Consider a geodesic γ in the class $[w] = [abab^{-1}]$. Let

$$M = \theta^{-1}(w) = ABAB^{-1} = \begin{pmatrix} 4 & 3 \\ 9 & 7 \end{pmatrix}$$

and note that M fixes an axis $\tilde{\gamma}$ in $\mathbf{H}$ with feet $p_1, p_2 = (-1 \pm \sqrt{13})/6$. Let $\mathcal{S} = [z_0, Mz_0]$ be the fundamental segment of $\tilde{\gamma}$, where z_0 is the point of intersection between $\tilde{\gamma}$ and $\mathcal{D}$ closest to the repulsive fixed point of $\tilde{\gamma}$. We know that γ is covered by and has the same orientation as the projection of $\mathcal{S}$ onto $\mathbf{T}$. Thus, it is sufficient to examine the lifts of $\mathcal{S}$ into $\mathcal{D}$. In particular, we can verify the number of self-intersections of γ by computing the number of self-intersections of the lifts of $\mathcal{S}$ into $\mathcal{D}$.

Note that $\mathcal{S}$ intersects as many copies of $\mathcal{D}$ as there are letters in M. We want to map the intersection of $\mathcal{S}$ with each of these squares back into $\mathcal{D}$. Thus, we apply successive portions of M^{-1} to $\mathcal{S}$. Because each of these lifts of $\mathcal{S}$ is a segment of a lift of $\tilde{\gamma}$, it is sufficient to examine the lifts of $\tilde{\gamma}$ which intersect $\mathcal{D}$. To compute these lifts, first compute

$$M^{-1} = BA^{-1}B^{-1}A^{-1}$$

and let $\mathcal{M}_k$ be the partial words formed by considering the last k generators in M^{-1}. These operations transform the original feet p_1 and p_2 into the feet of a lift of $\tilde{\gamma}$. Since $\mathcal{M}_4 M z_0 = z_0$, we will get every shift of $\mathcal{S}$ into $\mathcal{D}$ by considering $\tilde{\gamma}$ and the lifts $[\mathcal{M}_i p_1, \mathcal{M}_i p_2]$ for $i = 1, 2, 3$. These lifts are illustrated in Figure 2.1. Because the intersections at opposite edges of the region are identified, we can see that the lifts cross one time within the fundamental region. This agrees with the fact that γ has one self-intersection.

We can perform a similar operation with any hyperbolic word, because the axis of any such word will always pass through $\mathcal{D}$ [C]. In Section 3, for example, we will use a similar technique to show that particular conjugacy classes contain curves with a minimum of three self-intersections.

3. Loops with Two Self-Intersections

Our goal in this section is to classify geodesics on $\mathbf{T}$ with two self-intersections. We considered three methods for doing so. One possible method we considered was to consider all compositions of a simple loop and a loop with one non-trivial self-intersection. This will yield all classes of loops with two non-trivial self-intersections. However, many of the cases that must be considered are redundant. Furthermore, it is impossible to continue by way of induction to classify loops with n non-trivial self-intersections, because there are at least two classes of loops with three non-trivial self-intersections that are not generated by any composition of a loop with two non-trivial self-intersections and a simple loop. These classes, $[g(abab^{-1}a^{-1}ba^{-1}b^{-1})]$ and $[g(aaba^{-1}b^{-1}b^{-1})]$, will be discussed later.

Another possible method is an extension of Crisp's techniques. Consider one non-trivial intersection on a loop. The loop is the composition of two subloops at this intersection. If the original loop has two non-trivial self-intersections, one of two cases will arise.

(i) Each subloop is in the conjugacy class of a simple loop, and the two subloops cross non-trivially exactly once away from the chosen intersection.

(ii) One of the subloops has one non-trivial self-intersection and the other is in the conjugacy class of a simple loop. The two subloops do not cross non-trivially away from the chosen intersection.

By examining all possible subcases of (i) and (ii), it should be possible to classify all loops on $\mathbf{T}$ with two non-trivial self-intersections. However, it is difficult to know what all of these subcases are. In particular, we must examine more than one representative of each conjugacy class of loops. For example, $aaba^{-1}b^{-1}$ is automorphic to $bab^{-1}a^{-1}bab^{-1}$, yet they yield loops in the classes $[aaaba^{-1}b^{-1}]$ and $[abab^{-1}a^{-1}bab^{-1}]$, respectively, when composed with the generator a. We will later show that the first class and the second class, which is automorphic to $[abab^{-1}aba^{-1}b^{-1}]$, are not automorphic.

Finally, we considered the composition of three simple loops, as justified by the following lemma.

Lemma. *Any loop l on $\mathbf{T}$ with k non-trivial self-intersections can be formed as the composition of $k+1$ simple loops, which intersect at only one point.*

Proof. Consider a loop l on $\mathbf{T}$ with k non-trivial self-intersections occurring at distinct points. Because four segments converge at each intersection and each segment connects two intersections, there are exactly $4k/2 = 2k$ total distinct segments connecting the self-intersections of l.

Now, pick some point p on $\mathbf{T}$. For each self-intersection of l, we can continuously deform l so that the intersection occurs at p, in effect collapsing a segment of l. In order to bring all intersections to one point, we perform $k-1$ collapsings, leaving $2k-(k-1) = k+1$ segments. □

This lemma shows that it is sufficient to consider the compositions of three simple loops in order to generate all loops with two non-trivial self-intersections. We continue by classifying these compositions.

Theorem 3.1. *The conjugacy class in $\pi_1(\mathbf{T})$ of the composition of three simple loops on $\mathbf{T}$ is one of*

(a) $[(aba^{-1}b^{-1})^3]$ *or* $[(bab^{-1}a^{-1})^3]$
(b) $[g(aaba^{-1}b^{-1}aba^{-1}b^{-1})]$
(c) $[g(aaba^{-1}b^{-1}b^{-1})]$
(d) $[g(aba^{-1}bab^{-1})]$
(e) $[g(aaaba^{-1}b^{-1})]$
(f) $[g(abab^{-1}aba^{-1}b^{-1})]$
(g) $[g(aaba^{-1}a^{-1}b^{-1})]$
(h) $[g(abab^{-1}a^{-1}ba^{-1}b^{-1})]$
(i) $[g(a^3)]$
(j) $[g(aabab^{-1})]$
(k) *a simple loop*
(l) *a loop with one non-trivial self-intersection*

for some $g \in Aut\ \pi_1(\mathbf{T})$.

Proof. Let $l = l_1l_2l_3$, where l_1, l_2 and l_3 are simple loops on $\mathbf{T}$. We can move any self-intersections in l to the origin of l by a continuous deformation. Without loss of generality, we can assume that the origin of l is the base point $\sigma(i)$ of $\pi_1(\mathbf{T})$. Let w, w_1, w_2 and w_3 be the homotopy classes of l, l_1, l_2 and l_3 respectively. The corresponding free homotopy classes are $[w]$, $[w_1]$, $[w_2]$ and $[w_3]$.

Note that $[w] = [w_1w_2w_3] = [w_3w_1w_2] = [w_2w_3w_1]$ since cyclic permutations are conjugate. Therefore, we need consider at most one even and one odd permutation (we need consider only one permutation when any two loops are homotopic).

By Theorem 2.1, each of l_1, l_2 and l_3 can be either the identity, a generator, or a loop around the puncture. Thus, consider these cases:

Case 1: $[w_1] = [\mathrm{Id}]$.

In this case, $w = [w_2w_3]$, the conjugacy class of either a simple loop or a loop with one non-trivial self-intersection.

Case 2: $[w_1], [w_2], [w_3] \in \{[aba^{-1}b^{-1}], [bab^{-1}a^{-1}]\}$.

We know that $[aba^{-1}b^{-1}]$ and $[bab^{-1}a^{-1}]$ are the free homotopy classes of simple loops around the puncture, and that they have opposite orientations. There are two possibilities. Either $w_1 = w_2 = w_3$ or two loops are homotopic and one is reversed (without loss of generality, let $w_1 = w_2 = w_3^{-1}$). In the first instance,

$$[w] = [w_1^3] = [(aba^{-1}b^{-1})^3] \quad \text{or} \quad [(bab^{-1}a^{-1})^3], \tag{a}$$

depending on the orientation. Otherwise, $[w] = [w_1w_2w_3] = [w_3^{-1}] = [aba^{-1}b^{-1}]$ or $[bab^{-1}a^{-1}]$.

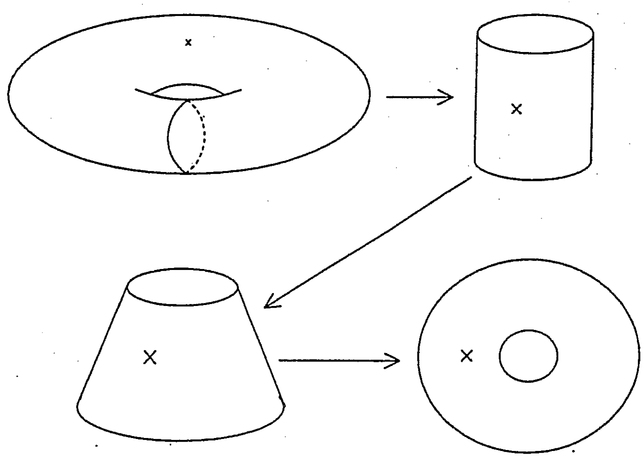

FIGURE 3.1. We can cut and deform $\mathbf{T}$ into a disk.

For the remaining cases, we make use of three methods from [C]. First, note that we can induce an automorphism between the w_i and words in a and b. Typically, we use $g(a) = w_j$ and $g(b) = w_k$ for $j \neq k$, but occasionally we use more complicated automorphisms in order to simplify the word. We also simplify words via cyclic reduction, which is actually conjugation with an appropriate word. In addition, note that every time we refer to an automorphism g, we actually mean *any* automorphism g. For example, $[g(a)]$ refers to any generator in $\pi_1(\mathbf{T})$. We allow this abuse in order to simplify the notation. The second method we adopt is the technique of cutting $\mathbf{T}$ along some loop l whose image in $\pi_1(\mathbf{T})$ is a generator to obtain a disk bounded by l containing the puncture and a hole also bounded by l. This operation is illustrated in Figure 3.1. Finally, we frequently introduce an additional path whose image in $\pi_1(\mathbf{T})$ forms a basis with the homotopy class of the loop bounding this disk.

Case 3: $[w_1] = [g(a)]$ and $[w_2], [w_3] \in \{[aba^{-1}b^{-1}], [bab^{-1}a^{-1}]\}$.

The loop l_1 does not separate $\mathbf{T}$, so cut $\mathbf{T}$ along l_1 to obtain a disk bounded by l_1 containing the puncture and a hole also bounded by l_1. Define a path l_4 such that w_1 and w_4 form a basis for $\pi_1(\mathbf{T})$. If l_2 and l_3 have opposite orientation, then $w_2 = w_3^{-1}$ and $[w] = [w_1] = [g(a)]$. Otherwise, $w_2 = w_3$ and the loop l_2l_3 is a double loop around

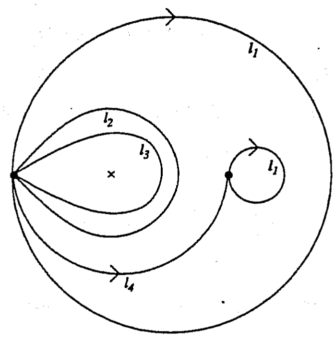

FIGURE 3.2. The torus $\mathbf{T}$ cut along l_1 in Case 3.

the puncture. We see from Figure 3.2 that l_2 and l_3 are homotopic to either $l_1 l_4 l_1^{-1} l_4^{-1}$ or $l_4 l_1 l_4^{-1} l_1^{-1}$, depending on their orientation. In the first case,

$$[w] = [w_1 w_2 w_3] = [w_1 w_1 w_4 w_1^{-1} w_4^{-1} w_1 w_4 w_1^{-1} w_4^{-1}] = [g(aaba^{-1}b^{-1}aba^{-1}b^{-1})], \quad \text{(b)}$$

and in the second $[w] = [w_1 w_2 w_3] = [w_1 w_4 w_1 w_4^{-1} w_1^{-1} w_4 w_1 w_4^{-1} w_1^{-1}] = [g(aaba^{-1}b^{-1})]$.

Case 4: $[w_1], [w_2] = [g(a)]$ and $[w_3] \in \{[aba^{-1}b^{-1}], [bab^{-1}a^{-1}]\}$.

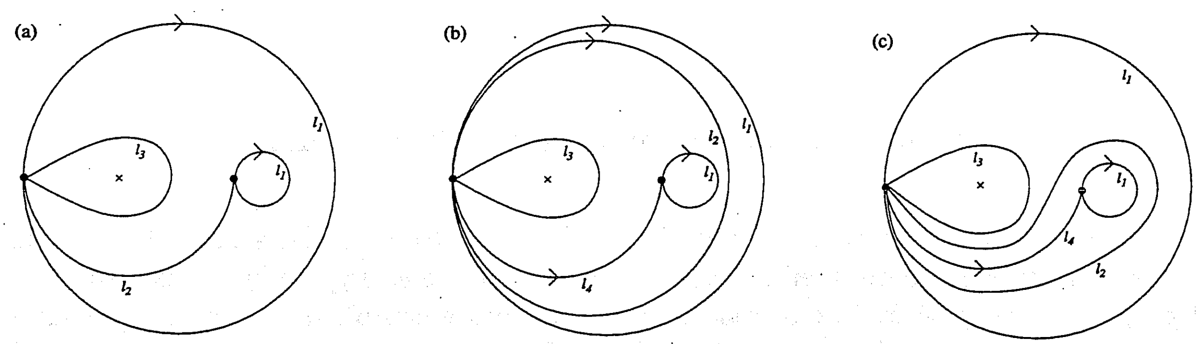

FIGURE 3.3. The torus $\mathbf{T}$ cut along l_1 in the first, second and third parts of Case 4.

Cut $\mathbf{T}$ along l_1. Since w_2 is a generator, but not necessarily equal to w_1, three cases arise.

First Case: l_2 is a loop such that w_1 and w_2 form a basis for $\pi_1(\mathbf{T})$. There are two possible orientations for each of l_2 and l_3, yielding four subcases each with two permutations. For each, see Figure 3.3(a).

(i) l_2 goes into the hole and l_3 is clockwise. Note that l_3 is homotopic to $l_1 l_2 l_1^{-1} l_2^{-1}$. Thus, $[w] = [w_1 w_2 w_3] = [w_1 w_2 w_1 w_2 w_1^{-1} w_2^{-1}] = [g(aaba^{-1}b^{-1})]$ or $[w] = [w_1 w_3 w_2] = [w_1 w_1 w_2 w_1^{-1} w_2^{-1} w_2] = [g(a)]$.

(ii) l_2 goes into the hole and l_3 is counterclockwise. Note that l_3 is homotopic to $l_2 l_1 l_2^{-1} l_1^{-1}$. Thus, $[w] = [w_1 w_2 w_3] = [w_1 w_2 w_2 w_1 w_2^{-1} w_1^{-1}] = [g(a)]$ or $[w] = [w_1 w_3 w_2] = [w_1 w_2 w_1 w_2^{-1} w_1^{-1} w_2] = [g(aaba^{-1}b^{-1})]$.

(iii) l_2 comes out of the hole and l_3 is clockwise. Note that l_3 is homotopic to $l_1 l_2^{-1} l_1^{-1} l_2$. Thus, $[w] = [w_1 w_2 w_3] = [w_1 w_2 w_1 w_2^{-1} w_1^{-1} w_2] = [g(aaba^{-1}b^{-1})]$ or

$$[w] = [w_1 w_3 w_2] = [w_1 w_1 w_2^{-1} w_1^{-1} w_2 w_2] = [g(aaba^{-1}b^{-1}b^{-1})] \tag{c}$$

(iv) l_2 comes out of the hole and l_3 is counterclockwise. Note that l_3 is homotopic to $l_2^{-1} l_1 l_2 l_1^{-1}$. Thus, $[w] = [w_1 w_2 w_3] = [w_1 w_2 w_2^{-1} w_1 w_2 w_1^{-1}] = [g(a)]$ or

$$[w] = [w_1 w_3 w_2] = [w_1 w_2^{-1} w_1 w_2 w_1^{-1} w_2] = [g(aba^{-1}bab^{-1})] \tag{d}$$

For the remaining two cases, define a path l_4, such that w_1 and w_4 form a basis for $\pi_1(\mathbf{T})$. Note that l_3 is homotopic to $l_1 l_4 l_1^{-1} l_4^{-1}$ or to $l_4 l_1 l_4^{-1} l_1^{-1}$, depending on its orientation.

Second Case: l_2 is homotopic to l_1 or l_1^{-1}. If l_2 is homotopic to l_1^{-1}, then $[w] = [w_1 w_2 w_3] = [w_3]$, a simple loop around the puncture. Otherwise, as in Figure 3.3(b), either

$$[w] = [w_1 w_2 w_3] = [w_1 w_1 w_1 w_4 w_1^{-1} w_4^{-1}] = [g(aaaba^{-1}b^{-1})], \tag{e}$$

or $[w] = [w_1 w_2 w_3] = [w_1 w_1 w_4 w_1 w_4^{-1} w_1^{-1}] = [g(abab^{-1})]$, depending on orientation.

Third Case: l_2 is a loop around the hole but not the puncture. From Figure 3.3(c), note that l_2 is homotopic to $l_4 l_1 l_4^{-1}$ or to $l_4 l_1^{-1} l_4^{-1}$. If l_2 is homotopic $l_4 l_1 l_4^{-1}$, with l_3 homotopic to $l_1 l_4 l_1^{-1} l_4^{-1}$, either

$$[w] = [w_1 w_2 w_3] = [w_1 w_4 w_1 w_4^{-1} w_1 w_4 w_1^{-1} w_4^{-1}] = [g(abab^{-1}aba^{-1}b^{-1})] \tag{f}$$

or $[w] = [w_1 w_3 w_2] = [w_1 w_1 w_4 w_1^{-1} w_4^{-1} w_4 w_1 w_4^{-1}] = [g(a^2)]$. With l_3 homotopic to $l_4 l_1 l_4^{-1} l_1^{-1}$, $[w] = [w_1 w_2 w_3] = [w_1 w_4 w_1 w_4^{-1} w_4 w_1 w_4^{-1} w_1^{-1}] = [g(a^2)]$, or $[w] = [w_1 w_3 w_2] = [w_1 w_4 w_1 w_4^{-1} w_1^{-1} w_4 w_1 w_4^{-1}] = [g(abab^{-1}aba^{-1}b^{-1})]$. Otherwise, l_2 is homotopic to $l_4 l_1^{-1} l_4^{-1}$. With l_3 homotopic to $l_1 l_4 l_1^{-1} l_4^{-1}$, $[w] = [w_1 w_2 w_3] = [w_1 w_4 w_1^{-1} w_4^{-1} w_1 w_4 w_1^{-1} w_4^{-1}] = [g((aba^{-1}b^{-1})^2)]$ or

$$[w] = [w_1 w_3 w_2] = [w_1 w_1 w_4 w_1^{-1} w_4^{-1} w_4 w_1^{-1} w_4^{-1}] = [g(aaba^{-1}a^{-1}b^{-1})]. \tag{g}$$

With l_3 homotopic to $l_4 l_1 l_4^{-1} l_1^{-1}$, $[w] = [w_1 w_2 w_3] = [w_1 w_4 w_1^{-1} w_4^{-1} w_4 w_1 w_4^{-1} w_1^{-1}] = [\mathrm{Id}]$ or

$$[w] = [w_1 w_3 w_2] = [w_1 w_4 w_1 w_4^{-1} w_1^{-1} w_4 w_1^{-1} w_4^{-1}] = [g(abab^{-1}a^{-1}ba^{-1}b^{-1})]. \tag{h}$$

Case 5: $[w_1], [w_2], [w_3] = [g(a)]$

Five cases arise.

First Case: l_2 and l_3 are each homotopic to l_1 or l_1^{-1}. Depending on the orientation of the loops, $[w] = [g(a)]$ or

$$[w] = [g(a^3)]. \tag{i}$$

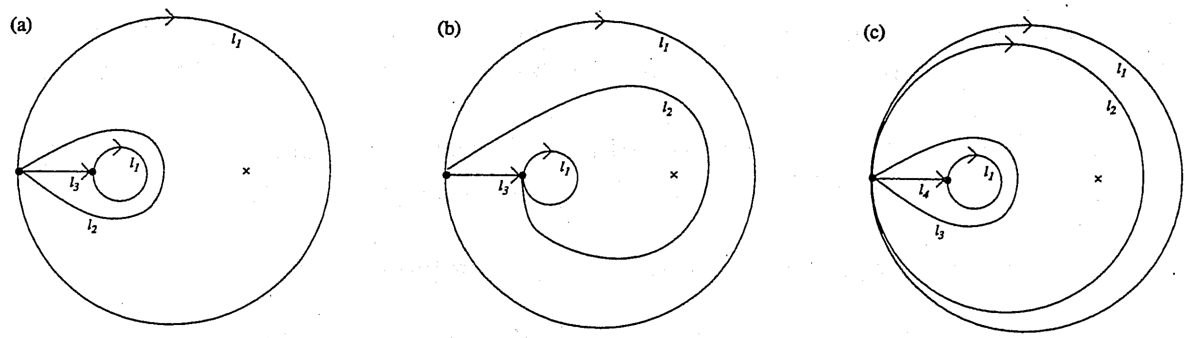

FIGURE 3.4. The torus $\mathbf{T}$ cut along l_1 in the second, third and fourth parts of Case 5.

Second Case: no loop is homotopic to any other loop or to its inverse, and two homotopy classes each form a basis for $\pi_1(\mathbf{T})$ with the third class, but not with each other. Assume that w_1 and w_3 form a basis for $\pi_1(\mathbf{T})$ and that w_2 and w_3 form a basis for $\pi_1(\mathbf{T})$. If we fix the orientation of l_1, a change in orientation of l_3 is realizable by an automorphism. Therefore, we also fix the orientation of l_3. Necessarily, w_1 and w_2 are conjugate. Cut $\mathbf{T}$ along l_1. From Figure 3.4(a), note that l_2 is homotopic to $l_3 l_1 l_3^{-1}$ or $l_3 l_1^{-1} l_3^{-1}$. When l_2 is homotopic to $l_3 l_1 l_3^{-1}$, $[w] = [w_1 w_2 w_3] = [w_1 w_3 w_1 w_3^{-1} w_3] = [g(a)]$ or

$$[w] = [w_1 w_3 w_2] = [w_1 w_3 w_3 w_1 w_3^{-1}] = [g(aaba^{-1}b)]. \tag{d'}$$

When l_2 is homotopic to $l_3 l_1^{-1} l_3^{-1}$, $[w] = [w_1 w_2 w_3] = [w_1 w_3 w_1^{-1} w_3^{-1} w_3] = [g(a)]$ or $[w] = [w_1 w_3 w_2] = [w_1 w_3 w_3 w_1^{-1} w_3^{-1}] = [g(aaba^{-1}b^{-1})]$.

Third Case: no loop is homotopic to any other loop or to its inverse, and any combination of two loops forms a basis for $\pi_1(\mathbf{T})$. If we assume w_1 and w_3 form a basis for $\pi_1(\mathbf{T})$, l_2 can be homotopic to $l_1 l_3$, $l_3 l_1$, $l_1 l_3^{-1}$, $l_3^{-1} l_1$ or to one of their inverses. We obtain similar results for all cases, so assume $w_2 = w_1 w_3$ or $w_3^{-1} w_1^{-1}$, as shown in Figure 3.4(b). In the first case, $[w] = [w_1 w_2 w_3] = [w_1 w_1 w_3 w_3] = [g(abab^{-1})]$ or $[w] = [w_1 w_3 w_2] = [w_1 w_3 w_1 w_3] = [g(a^2)]$. In the second, $[w] = [w_1 w_2 w_3] = [w_1 w_3^{-1} w_1^{-1} w_3] = [bab^{-1}a^{-1}]$ or $[w] = [w_1 w_3 w_2] = [w_1 w_3 w_3^{-1} w_1^{-1}] = [\mathrm{Id}]$.

For the remaining two cases, define a path l_4 such that w_1 and w_4 form a basis for $\pi_1(\mathbf{T})$.

Fourth Case: l_2 is homotopic to l_1 or l_1^{-1} and l_3 is a loop around the hole but not the puncture. If l_2 is homotopic to l_1^{-1} then $[w] = [w_3] = [g(a)]$. Otherwise, cut $\mathbf{T}$ along l_1. Note, from Figure 3.4(c), that l_3 is homotopic to $l_4 l_1 l_4^{-1}$ or $l_4 l_1^{-1} l_4^{-1}$. When l_3 is homotopic to $l_4 l_1 l_4^{-1}$,

$$[w] = [w_1 w_2 w_3] = [w_1 w_1 w_4 w_1 w_4^{-1}] = [g(aabab^{-1})]. \tag{j}$$

Otherwise, $[w] = [w_1 w_2 w_3] = [w_1 w_1 w_4 w_1^{-1} w_4^{-1}] = [g(aaba^{-1}b^{-1})]$.

Fifth Case: l_2 is homotopic to l_1 or l_1^{-1} and l_3 is a loop such that w_1 and w_3 form a basis for $\pi_1(\mathbf{T})$. If l_2 is homotopic to l_1^{-1} then $[w] = [w_3] = [g(a)]$. Otherwise, $[w] = [w_1 w_2 w_3] = [w_1 w_1 w_3] = [g(a)]$.

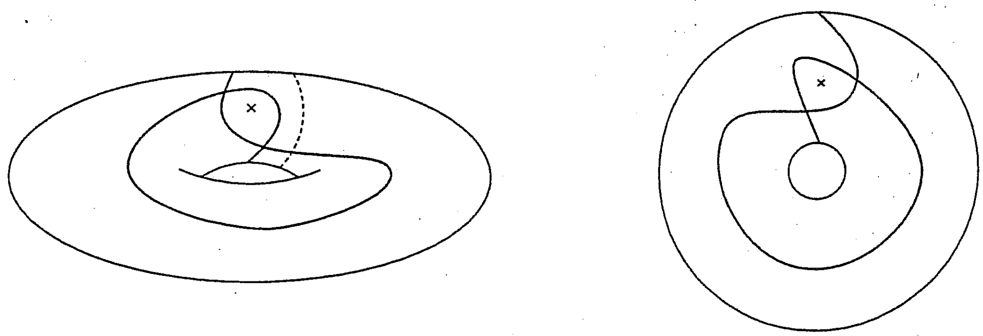

FIGURE 3.5. The punctured torus $\mathbf{T}$ with a loop in the free homotopy class $[aba^{-1}bab^{-1}]$ can be cut along the indicated generator and deformed into this disk.

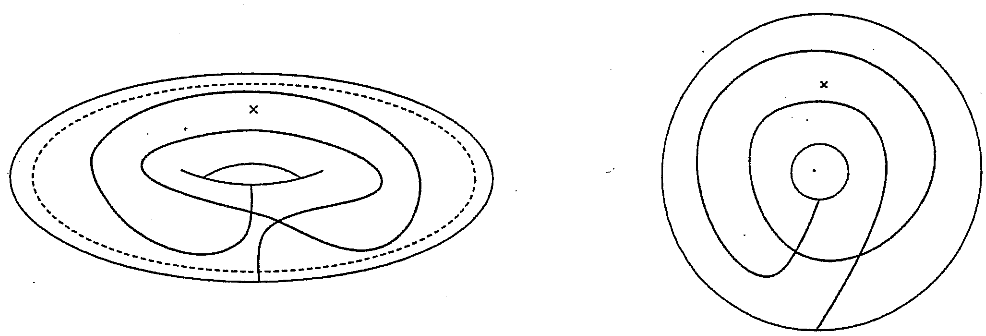

FIGURE 3.6. The punctured torus $\mathbf{T}$ with a loop in the free homotopy class $[bba^{-1}b^{-1}a^{-1}]$ can be cut along the indicated generator and deformed into this disk.

Unfortunately, this technique will generate loops that are automorphic, but are not obviously so. Consider the loop $aba^{-1}bab^{-1}$ in the conjugacy class $[g(aba^{-1}bab^{-1})]$. We can cut the torus along the indicated dashed line and deform it into a disk. This is illustrated in Figure 3.5. The loop $bba^{-1}b^{-1}a^{-1}$ in the conjugacy class $[g(aaba^{-1}b)]$ can be cut and deformed into a disk as shown in Figure 3.6. We can go between the two disks by rotating the inside edge of the first by one half twist clockwise and the outside edge one half twist counterclockwise. Together, these are a full 360° twist, a homeomorphism on $\mathbf{T}$. We know that there is a correspondence between automorphisms on $\pi_1(\mathbf{T})$ and homotopy classes of homeomorphisms on $\mathbf{T}$ [N]. Therefore, there is an automorphism g_0 on $\pi_1(\mathbf{T})$ such that $g_0(aba^{-1}bab^{-1}) = bba^{-1}b^{-1}a^{-1}$. In fact, we can deduce the proper automorphism, $a \mapsto a^{-1}$ and $b \mapsto a^{-1}b$, from the nature of the homeomorphism. Thus, the original free homotopy classes are automorphic.

This occurs because of the way we collapse segments to move all intersections to one point. We cannot collapse segments that connect an intersection to itself. All loops with two non-trivial self-intersections and no trivial self-intersections have four segments. In all but one class, two segments are non-collapsible and collapsing each of the remaining two segments yields symmetric cases. In the above class, however, all four segments are collapsible and there are two pairs of symmetric cases. Thus, this class appears twice, in apparently different cases.

From Birman and Series, we know which of these composed loops are simple, and from Crisp, we know which have one non-trivial self-intersection. We will list the conjugacy classes of the remaining loops, without attempting to classify them at this time. □

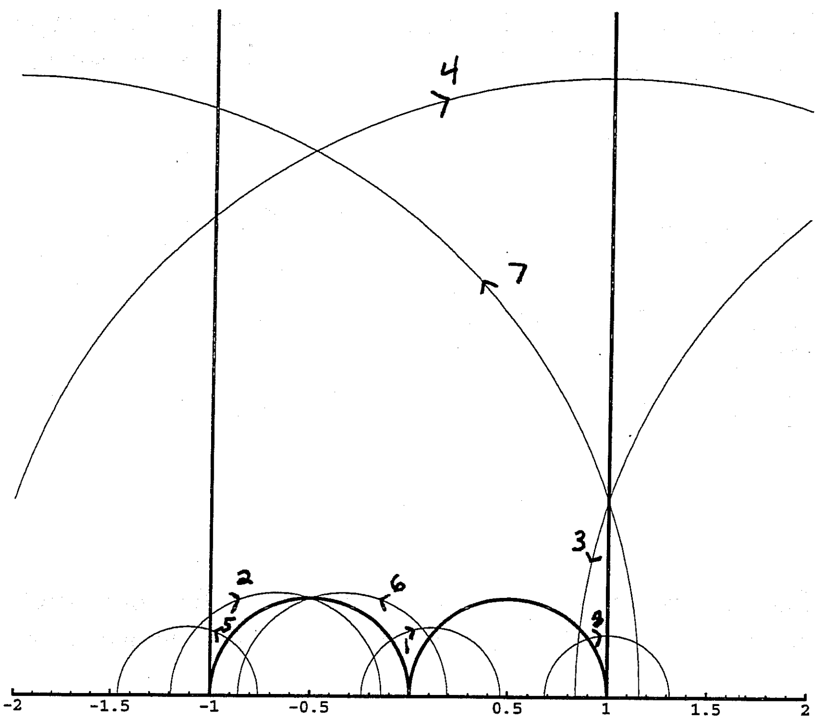

FIGURE 3.7. Lifts in $\mathbf{H}$ of the axis of $M = ABAB^{-1}A^{-1}BA^{-1}B^{-1}$.

Now, we will demonstrate that two of these classes contain loops with three non-trivial self-intersections. First, consider the conjugacy class $[g(abab^{-1}a^{-1}ba^{-1}b^{-1})]$. Because geodesics always realize the minimal number of self-intersections in their class, consider a geodesic γ in this class on $\mathbf{T}$. Compute

$$M = ABAB^{-1}A^{-1}BA^{-1}B^{-1} = \begin{pmatrix} 25 & 6 \\ 54 & 13 \end{pmatrix},$$

and note that it fixes an axis $\tilde{\gamma}$ in $\mathbf{H}$ with feet $p_1, p_2 = (1 \pm \sqrt{10})/9$. Let $\mathcal{S} = [z_0, Mz_0]$ be the fundamental segment of $\tilde{\gamma}$, where z_0 is the point of intersection between $\tilde{\gamma}$ and $\mathcal{D}$

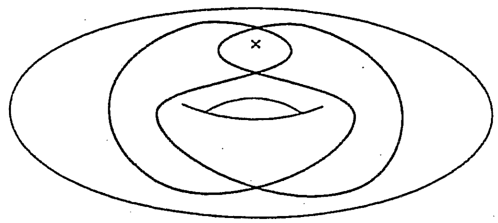

FIGURE 3.8. A loop on $\mathbf{T}$ in the free homotopy class $[abab^{-1}a^{-1}ba^{-1}b^{-1}]$ has three non-trivial self-intersections.

closest to the repulsive fixed point of $\tilde{\gamma}$. To find the number of self-intersections of γ, it is sufficient to compute the number of self-intersections of the lifts of $\mathcal{S}$ into $\mathcal{D}$.

Each of these lifts is a segment of a lift of $\tilde{\gamma}$. To compute the lifts of $\tilde{\gamma}$, first compute

$$M^{-1} = BAB^{-1}ABA^{-1}B^{-1}A^{-1},$$

and let $\mathcal{M}_k$ be the partial words formed by considering the last k generators in M^{-1}. These operations transform the original feet p_1 and p_2 into the feet of a lift of $\tilde{\gamma}$. Since $\mathcal{M}_8 M z_0 = z_0$, we will get every shift of $\mathcal{S}$ into $\mathcal{D}$ by considering $\tilde{\gamma}$ and the lifts $[\mathcal{M}_i p_1, \mathcal{M}_i p_2]$ for $i = 1, 2, \ldots, 7$.

Figure 3.7 illustrates that these lifts cross three times within the fundamental region (the intersections at opposite edges of the region are identified). Thus, γ is a closed geodesic with three non-trivial self-intersections as shown in Figure 3.8. It follows that $[g(abab^{-1}a^{-1}ba^{-1}b^{-1})]$ has three non-trivial self-intersections for any g because automorphisms on $\pi_1(\mathbf{T})$ preserve the number of intersections [C].

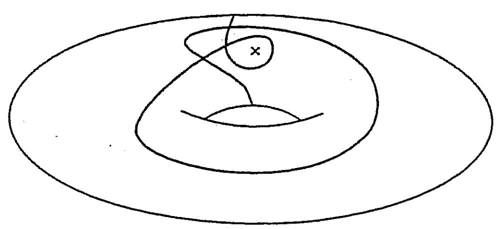

FIGURE 3.10. A loop on $\mathbf{T}$ in the free homotopy class $[aaba^{-1}b^{-1}b^{-1}]$ has three non-trivial self-intersections.

Now, consider a geodesic γ on $\mathbf{T}$ in the conjugacy class $[g(aaba^{-1}b^{-1}b^{-1})]$ and compute

$$M = AABA^{-1}B^{-1}B^{-1} = \begin{pmatrix} -15 & -8 \\ -28 & -15 \end{pmatrix}$$

and note that it fixes an axis $\tilde{\gamma}$ in $\mathbf{H}$ with feet $p_1, p_2 = \pm\sqrt{2/7}$. Let $\mathcal{S} = [z_0, Mz_0]$ be the fundamental segment of $\tilde{\gamma}$, where z_0 is the point of intersection between $\tilde{\gamma}$ and $\mathcal{D}$ closest to

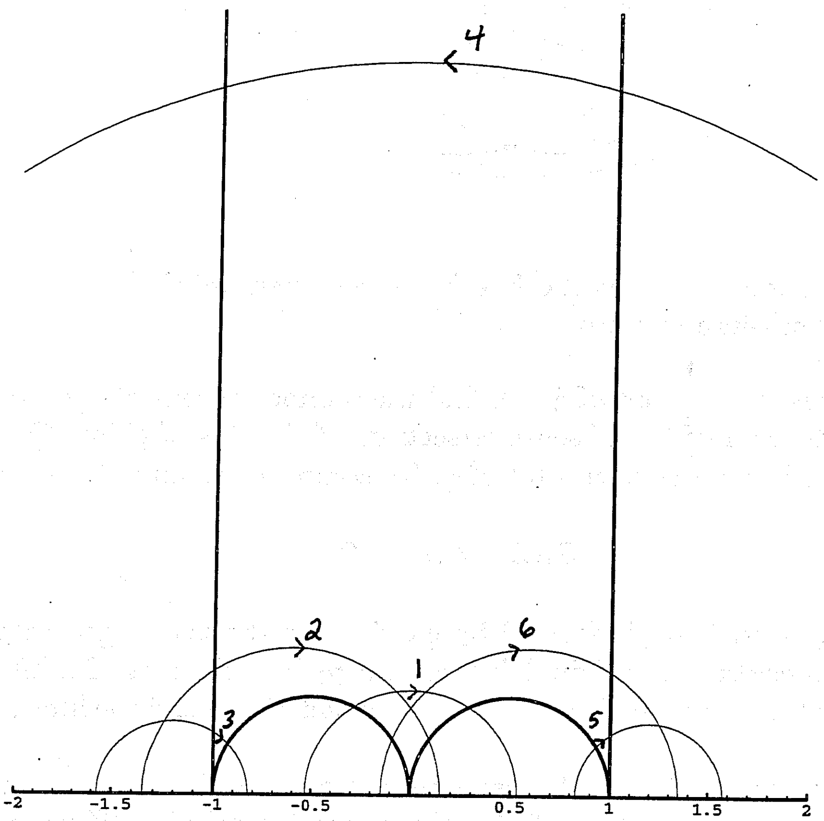

FIGURE 3.9. Lifts in $\mathbf{H}$ of the axis of $M = AABA^{-1}B^{-1}B^{-1}$.

the repulsive fixed point of $\tilde{\gamma}$. To find the number of self-intersections of γ, it is sufficient to compute the number of self-intersections of the lifts of $\mathcal{S}$ into $\mathcal{D}$.

Each of these lifts is a segment of a lift of $\tilde{\gamma}$. To compute the lifts of $\tilde{\gamma}$, first compute

$$M^{-1} = BBAB^{-1}A^{-1}A^{-1}$$

and let $\mathcal{M}_k$ be the partial words formed by considering the last k generators in M^{-1}. These operations transform the original feet p_1 and p_2 into the feet of a lift of $\tilde{\gamma}$. Since $\mathcal{M}_6 M z_0 = z_0$, we will get every shift of $\mathcal{S}$ into $\mathcal{D}$ by considering $\tilde{\gamma}$ and the lifts $[\mathcal{M}_i p_1, \mathcal{M}_i p_2]$ for $i = 1, 2, \ldots, 5$. Figure 3.6 illustrates that these lifts cross three times within the fundamental region. Thus, γ is a closed geodesic with three non-trivial self-intersections as shown in Figure 3.10. It follows that $[g(aaba^{-1}b^{-1}b^{-1})]$ has three non-trivial self-intersections for any g.

We have completed the classification of the compositions of three simple loops and identified that two conjugacy classes contain loops with three non-trivial self-intersections. Now, we can identify which of the remaining classes contain loops with two non-trivial self-intersections.

Theorem 3.2. *The conjugacy class in $\pi_1(\mathbf{T})$ of a loop on* $\mathbf{T}$ *with two non-trivial self-intersections is one of*

(a) $[(aba^{-1}b^{-1})^3]$ *or* $[(bab^{-1}a^{-1})^3]$
(b) $[g(aaba^{-1}b^{-1}aba^{-1}b^{-1})]$
(c) $[g(aba^{-1}bab^{-1})]$
(d) $[g(aaaba^{-1}b^{-1})]$
(e) $[g(abab^{-1}aba^{-1}b^{-1})]$
(f) $[g(aaba^{-1}a^{-1}b^{-1})]$
(g) $[g(a^3)]$
(h) $[g(aabab^{-1})]$

for some $g \in Aut\ \pi_1(\mathbf{T})$*. Conversely, each of these conjugacy classes contains such a loop.*

Proof. By the Lemma, it is sufficient to consider the composition of three simple loops. We can now apply Theorem 3.1 and eliminate simple loops, loops with single non-trivial self-intersection and the previously discussed classes of loops with three non-trivial self-intersections.

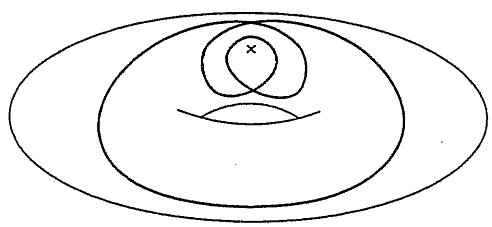

FIGURE 3.11. A loop on $\mathbf{T}$ in the free homotopy class $[aaba^{-1}b^{-1}aba^{-1}b^{-1}]$.

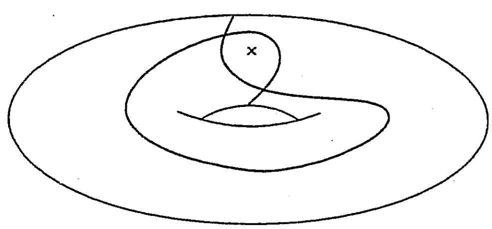

FIGURE 3.12. A loop on $\mathbf{T}$ in the free homotopy class $[aba^{-1}bab^{-1}]$.

The only candidates remaining are (a)–(h) above. It is sufficient to demonstrate a closed loop with two non-trivial self-intersections for each class in the case where g is the identity, because automorphisms on $\pi_1(\mathbf{T})$ preserve the number of self-intersections. For $[a^3]$, the obvious choice is a tripled generator, and for each of $[(aba^{-1}b^{-1})^3]$ and $[(bab^{-1}a^{-1})^3]$, the obvious choice is a triple loop around the puncture. This disposes of classes (a) and (g). For the remaining cases, see Figures 3.11 – 3.16. $\square$

We have now demonstrated a complete classification of loops with two non-trivial self-intersections. Now, we will demonstrate that all of these classes are distinct. First, note

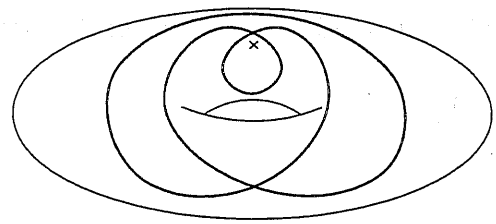

FIGURE 3.13. A loop on $\mathbf{T}$ in the free homotopy class $[aaaba^{-1}b^{-1}]$.

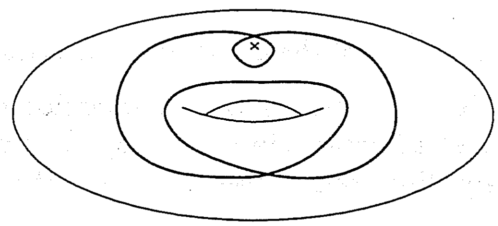

FIGURE 3.14. A loop on $\mathbf{T}$ in the free homotopy class $[g(abab^{-1}aba^{-1}b^{-1})]$.

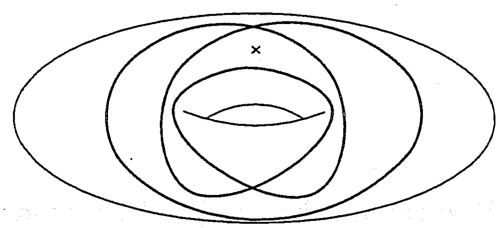

FIGURE 3.15. A loop on $\mathbf{T}$ in the free homotopy class $[aabab^{-1}]$.

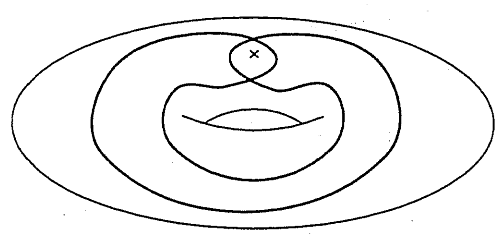

FIGURE 3.16. A loop on $\mathbf{T}$ in the free homotopy class $[aaba^{-1}a^{-1}b^{-1}]$.

that $[g(a^3)]$, $[(aba^{-1}b^{-1})^3]$ and $[(bab^{-1}a^{-1})^3]$ are distinct from the other classes because they are non-primitive, and are distinct from each other because $[a^3]$ corresponds to a hyperbolic word in $\mathbf{H}$, while the others correspond to parabolic words. Next, note that two loops on $\mathbf{T}$ cannot be in automorphic conjugacy classes in $\pi_1(\mathbf{T})$ if they are not in automorphic conjugacy classes on the normal torus T. When we remove the puncture from $\mathbf{T}$, loops can be deformed as follows:

(i) $[aaba^{-1}a^{-1}b^{-1}]$ is in $[\mathrm{Id}]$.
(ii) $[aaba^{-1}b^{-1}aba^{-1}b^{-1}]$ and $[aba^{-1}bab^{-1}]$ are in $[g(a)]$
(iii) $[aaaba^{-1}b^{-1}]$ and $[abab^{-1}aba^{-1}b^{-1}]$ are in $[g(a^2)]$
(iv) $[aabab^{-1}]$ is in $[g(a^3)]$

for some $g \in Aut\ \pi_1(T)$. This enables us to distinguish all but two pairs of classes.

Suppose two conjugacy classes $[w_1]$ and $[w_2]$ are automorphic. Each class contains a unique geodesic. Specifically, let γ_1 represent $[w_1]$ and γ_2 represent $[w_2]$. We know that there exists a homeomorphism that takes γ_1 to a loop in $[w_2]$ and takes γ_2 to a loop in $[w_1]$. We are fairly sure that the following can be proved.

Assumption. *If $[w_1]$ and $[w_2]$ are automorphic classes in $\pi_1(\mathbf{T})$ with geodesics γ_1 and γ_2 respectively, there is a homeomorphism $h\colon \mathbf{T} \to \mathbf{T}$ that takes γ_1 to γ_2.*

If this assumption is true, h can be restricted to be a homeomorphism from a subset of $\mathbf{T}$ to its image. Thus, $\mathbf{T} \setminus \{\gamma_1\}$ is homeomorphic to $\mathbf{T} \setminus \{\gamma_2\}$. In particular, this implies that the number of regions into which a geodesic separates $\mathbf{T}$ is invariant under automorphism. This allows us to easily distinguish $[aaba^{-1}b^{-1}aba^{-1}b^{-1}]$ from $[aba^{-1}bab^{-1}]$.

This homeomorphism between $\mathbf{T} \setminus \{\gamma_1\}$ and $\mathbf{T} \setminus \{\gamma_2\}$ also preserves certain topological properties that enable us to distinguish $[w_1] = [aaaba^{-1}b^{-1}]$ and $[w_2] = [abab^{-1}aba^{-1}b^{-1}]$. Since the geodesics in each of these classes separate $\mathbf{T}$ into three regions, we need to further examine these topological properties. Suppose there is some homeomorphism h that maps γ_1 to γ_2. In Figures 3.17 and 3.18, note that the shaded region surrounds a region containing the puncture. Thus h must map the shaded region in Figure 3.17 to the shaded region in Figure 3.18. We can cut along γ_1 and γ_2 to separate these regions. The first region is a disk with two non-manifold points, or "pinches." The second is an annulus with two such points. These points are formed by identifying adjacent corresponding heavy dots shown in each figure. But no homeomorphism can take a disk to an annulus. Therefore no such homeomorphism h can exist, and there can be no automorphism from $[w_1]$ to $[w_2]$.

It remains only to identify which of these classes of loops contain geodesics. We will show that all closed geodesics with two self-intersections on $\mathbf{T}$ can be classified by their conjugacy classes in Γ'. This proof parallels Theorem 3.2 in [C], but it is included for completeness.

Theorem 3.3. *A closed geodesic on $\mathbf{T}$ has two self-intersections if and only if it is defined by a conjugacy class in Γ' of one of the following forms*

(a) $[G(AABA^{-1}B^{-1}ABA^{-1}B^{-1})]$
(b) $[G(AAABA^{-1}B^{-1})]$
(c) $[G(ABAB^{-1}ABA^{-1}B^{-1})]$
(d) $[G(ABA^{-1}BAB^{-1})]$
(e) $[G(AABA^{-1}A^{-1}B^{-1})]$
(f) $[G(AABAB^{-1})]$

for some $G \in Aut\ \Gamma'$.

Proof. Suppose a closed geodesic on $\mathbf{T}$ has two self-intersections. We know that closed geodesics on $\mathbf{T}$ are primitive, hyperbolic and realize the minimum number of self-

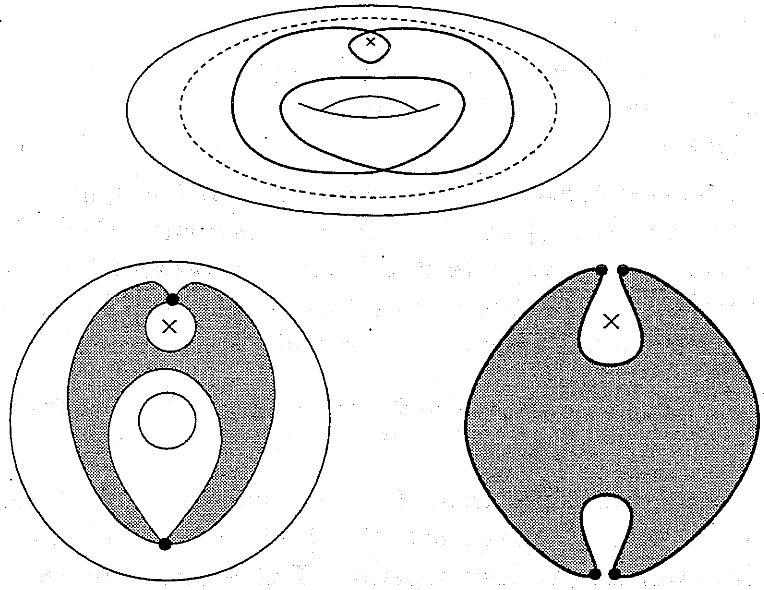

FIGURE 3.17. The punctured torus **T** with a loop in the free homotopy class $[abab^{-1}aba^{-1}b^{-1}]$ can be cut along the indicated generator and deformed into this disk. Then, the shaded region can be separated from the remainder of the disk.

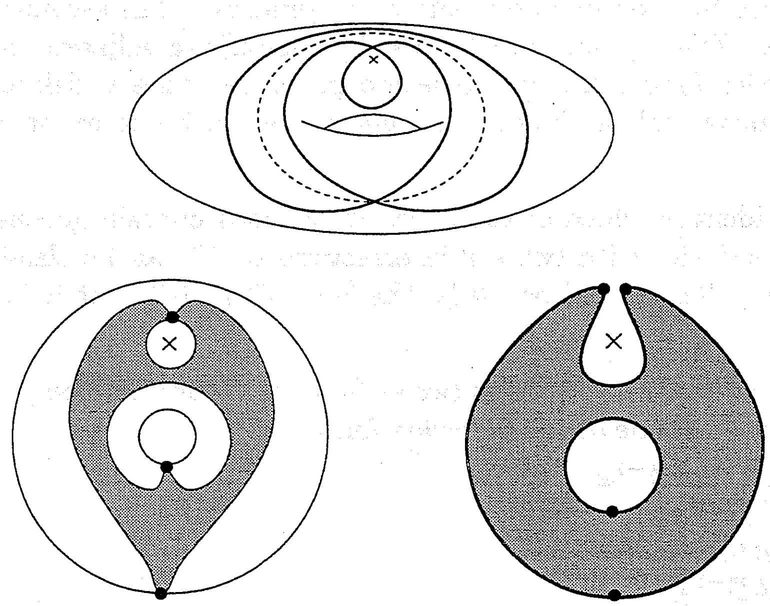

FIGURE 3.18. The punctured torus **T** with a loop in the free homotopy class $[aaaba^{-1}b^{-1}]$ can be cut along the indicated generator and deformed into this disk. Then, the shaded region can be separated from the remainder of the disk.

intersections for their free homotopy classes. We can apply Theorem 3.2 and preclude the non-primitive classes $[(aba^{-1}b^{-1})^3]$, $[(bab^{-1}a^{-1})^3]$ and $[g(a^3)]$. Only $[(aba^{-1}b^{-1})^3]$ and $[(bab^{-1}a^{-1})^3]$ are parabolic. We conclude that the conjugacy classes in $\pi_1(\mathbf{T})$ of closed geodesics with two self-intersections are of the remaining forms. Because we have an isomorphism θ between Γ' and $\pi_1(\mathbf{T})$, we can consider the image of each of these classes in Γ'. Thus, each conjugacy class in Γ' which defines a geodesic on $\mathbf{T}$ with two self-intersections is of one of the forms (a)–(f) above.

Suppose a loop on $\mathbf{T}$ is defined by one of the forms (a)–(f) in Γ'. It is clear that these forms are primitive and hyperbolic; thus their corresponding free homotopy classes in $\pi_1(\mathbf{T})$ contain closed geodesics. By Theorem 3.2, these free homotopy classes also contain loops with two non-trivial self-intersections. Geodesics realize the minimum number of self-intersections in their free homotopy classes. Therefore, these free homotopy classes must contain closed geodesics with two self-intersections. □

4. Related Results

Generating Loops with More Intersections.

In order to generate all classes of loops with two non-trivial self-intersections one must examine significantly more cases than are needed to generate loops with one such intersection. In addition, these cases are noticeably more complex, and non-obvious automorphic classes occur. There is every reason to suspect that this trend towards difficulty continues as the number of intersections increases. Thus, considering the composition of $k+1$ loops is not a practical method for generating loops with k non-trivial self-intersections if k is not very small. We will therefore examine an alternate method of generating all automorphism classes of loops with k self-intersections.

Every such class of loops is represented by some set of cyclically-reduced words that are not made shorter by any automorphism in $\pi_1(\mathbf{T})$. At least one of these minimal words begins with a and has b as its first non-a letter (if any). We claim, without proof, that one of these words begins with a longer series of a than any other, and call this word a "standard minimal word."

For a given length, there are probably not very many standard minimal words. For example, Table 4.1 lists standard minimal words of length one through five. If $aaabb$ looks unfamiliar, note that it is automorphic to $aba^{-1}bab^{-1}$.

Table 4.1. Standard minimal words of length 1, ... ,5.

Length	Words
1	a
2	a^2
3	a^3
4	a^4, $abab^{-1}$, $aba^{-1}b^{-1}$
5	a^5, $aaabb$, $aabab^{-1}$, $aaba^{-1}b^{-1}$

Conjecture 4.1. *The longest standard minimal word in $\pi_1(\mathbf{T})$ that specifies a class containing a loop with k non-trivial self-intersections is $(aba^{-1}b^{-1})^{k+1}$.*

If this is true, note that the shortest standard minimal word representing loops with k non-trivial self-intersections is a^{k+1}. Neither a^{k+1} nor $(aba^{-1}b^{-1})^{k+1}$ are geodesics, so we have demonstrated how to find geodesics on $\mathbf{T}$ with k non-trivial self-intersections.

Corollary. *In order to find all geodesics on $\mathbf{T}$ with k self-intersections, it is sufficient to examine all standard minimal words in $\pi_1(\mathbf{T})$ of length between $4k$ and $k+1$, exclusive.*

Thus, we need only find all standard minimal words of appropriate length, graph lifts into the fundamental region $\mathcal{D}$, and and count their intersections as we did in Sections 2 and 3. Unfortunately, no systematic way to generate all these words is apparent. In particular, we need to be able to quickly tell if two given words are automorphic. In Section 3, automorphisms between the classes of loops were always relatively simple, but were not always easy to find. There are seventy-two different automorphisms that either map a to a two-letter word and b to a one-letter word, b to a two-letter word and a to a one-letter word, or both a and b to one-letter words. Sometimes the appropriate automorphism was obvious. For example ab is clearly a generator, so $[ab] = [g(a)]$ for some $g \in \text{Aut } \pi_1(\mathbf{T})$. Others were less obvious and grew out of a geometric argument. Finding an automorphism between two large words can be difficult. In addition, the topological arguments that two words are distinct will become more complicated as the number of intersections increases.

Connections to the Markoff Spectrum.

Once Crisp had classified the geodesics on $\mathbf{T}$ with one non-trivial self-intersection, he used these results to find isolated values in the Markoff spectrum. Before we comment on similar conjectures about the relation between the Markoff spectrum and the geodesics we have classified, a brief introduction is needed. A more detailed introduction to the theory can be found in [CF], but we paraphrase from Crisp.

The Markoff spectrum is a subset of the interval $[\sqrt{5}, \infty)$ and is the set of values which arises from the study of the normalized minima of real indefinite binary quadratic forms,

$$f(x,y) = \alpha x^2 + \beta xy + \gamma y^2, \qquad \alpha, \beta, \gamma \in \mathbb{R}$$

with discriminant $d(f) = \beta^2 - 4\alpha\gamma > 0$. We define

$$m(f) = \inf \{|f(x,y)| : (x,y) \in \mathbb{Z} \times \mathbb{Z},\ (x,y) \neq (0,0)\}$$

and

$$M(f) = \sqrt{d(f)}/m(f).$$

The quantity $M(f)$ is called the Markoff value of the form f, and the set of Markoff values taken over all possible real indefinite binary quadratic forms is called the Markoff spectrum. Some facts are known about the structure of the Markoff spectrum. The portion of the spectrum which is less than 3 is a discrete set of values which converges to 3. It is also known that the spectrum contains a maximal interval of the form $[\nu, \infty)$ where ν is less than $\sqrt{21} \approx 4.582$. This is known as Hall's ray. Harvey Cohn discovered a connection between the properties of the Markoff spectrum and the behavior of geodesics on $\mathbf{T}$. If we associate a Markoff form f with the geodesic γ in $\mathbf{H}$ whose feet are the roots of f and projecting γ onto $\mathbf{T}$, these Markoff forms correspond to the closed geodesics.

Once Crisp identified the conjugacy classes in $\pi_1(\mathbf{T})$ of loops with one non-trivial self-intersection, he identified which of these classes contained geodesics. Of the five classes, two contained geodesics: $[g(abab^{-1})]$ and $[g(aaba^{-1}b^{-1})]$. Because the geodesics in the second class contain a subloop which bounds a disk containing the puncture, their Markoff values lie in Hall's ray. Notice that the subloop bounding such a disk is in the conjugacy class of $[aba^{-1}b^{-1}]$ or its inverse. In Γ', this corresponds to either $[ABA^{-1}B^{-1}]$ or its inverse. Crisp proves that this implies the geodesic corresponding to this class has a lift to $\mathbf{H}$ which has a diameter greater than 6, which is greater than ν. Thus the geodesics in $[g(aaba^{-1}b^{-1})]$ lie in Hall's ray. However, Crisp was able to calculate the Markoff values for the geodesics in the first class and conjectured that these values would be isolated points in the Markoff spectrum.

We have identified the conjugacy classes in $\pi_1(\mathbf{T})$ which contain loops with two non-trivial self-intersections and we have shown that six of these classes contain geodesics. Of these six classes, it is clear that at least three of them will have corresponding Markoff values which lie in Hall's ray. Specifically, loops in each of the automorphism classes $[g(aaba^{-1}b^{-1}aba^{-1}b^{-1})]$, $[g(aaaba^{-1}b^{-1})]$ and $[g(abab^{-1}aba^{-1}b^{-1})]$ each contain a subloop which bounds a disk which contains the puncture. Because we know that any automorphism of $\pi_1(\mathbf{T})$ can be induced by a homeomorphism of $\mathbf{T}$, the geodesics for these classes will also contain a subloop which bounds a disk containing the puncture. For this reason, we can apply Crisp's argument and conclude that the associated Markoff values will lie in Hall's ray.

When we examine the remaining three classes of loops, the classes $[g(aba^{-1}bab^{-1})]$ and $[g(aaba^{-1}a^{-1}b^{-1})]$ bound a region which is a disk containing the puncture, but it is not bounded by a subloop. It is unclear whether the associated Markoff values to the geodesics in these classes will lie in Hall's ray. The loops in the class $[g(aabab^{-1})]$ do not contain any such regions, so we make the following conjecture.

Conjecture 4.2. *The geodesics in the class $[g(aabab^{-1})]$ for all $g \in$ Aut $\pi_1(\mathbf{T})$ have associated Markoff which are isolated in the Markoff spectrum.*

5. Conclusion

We have classified all six distinct classes of hyperbolic geodesics with two non-trivial self-intersections on the punctured torus. We compared three methods for doing this, and decided to consider the composition of three simple loops. Unfortunately, while two of these methods have obvious generalizations, the number of cases that must be considered grows incredibly with the number of intersections.

Nevertheless, we have conjectured techniques to generate all classes of loops with an arbitrary number of self-intersections. In addition, we discussed the Markoff value associated to a geodesic, and identified three classes with associated values in Hall's ray. We also conjecture that geodesics in one automorphism class have associated Markoff values which are isolated.

References

[BS] Joan S. Birman and Caroline Series, *An Algorithm for Simple Curves on Surfaces*, J. London Math Soc. (2) **29** (1984), 331–342.

[C] David J. Crisp, *The Markoff Spectrum and Geodesics on the Punctured Torus*, Ph.D. thesis, Univ. Adelaide, 1993.

[CM] D. Crisp and W. Moran, *Single self-intersection geodesics and the Markoff spectrum*, The Markoff Spectrum, Diophantine Analysis and Analytic Number Theory (A. Pollington and W. Moran, eds.), Marcel Dekker, New York, 1993, pp. 83–94.

[CMZ] M. Cohen, W. Metzler, and A. Zimmermann, *What Does a Basis of $F(a, b)$ Look Like?*, Math. Ann. **257** (1981), 435–445.

[CF] T.W. Cusick and M.E. Flahive, *The Markoff and Lagrange spectra*, Mathematical Surveys and Monographs **30**, American Mathematical Society, Providence, Rhode Island, 1989.

[M] James R. Munkres, *Topology: A First Course*, Prentice Hall, Englewood Cliffs, New Jersey, 1975.

[N] J. Nielsen, *Die Isomorphismen der allgemeinen unendlichen Gruppe mit zwei Erzeugenden*, Math. Ann. **78** (1918), 385–397.

[Se1] Caroline Series, *The Geometry of Markoff Numbers*, Mathematical Intelligencer **7** (1985), 20–29.

[Se2] ———, *Non-Euclidean Geometry, Continued Fractions, and Ergodic Theory*, Mathematical Intelligencer **4** (1982), 24–31.

[St] John Stillwell, *Classical Topology and Combinatorial Group Theory*, Second Edition, Springer-Verlag, New York, 1993.

Department of Mathematics, Oregon State University, Corvallis OR 97331

E-mail: dziadosz@math.orst.edu, tinsel@math.orst.edu, wilesp@math.orst.edu