PARAMETRIZING SIMPLE CLOSED GEODESY ON $\Gamma^3\backslash\mathscr{H}$

THOMAS A. SCHMIDT and MARK SHEINGORN

Abstract

We exhibit a canonical geometric pairing of the simple closed curves of the degree three cover of the modular surface, $\Gamma^3\backslash\mathscr{H}$, with the proper single self-intersecting geodesics of Crisp and Moran. This leads to a pairing of fundamental domains for Γ^3 with Markoff triples.

The routes of the simple closed geodesics are directly related to the above. We give two parametrizations of these. Combining with work of Cohn, we achieve a listing of all simple closed geodesics of length within any bounded interval. Our method is direct, avoiding the determination of geodesic lengths below the chosen lower bound.

2000 *Mathematics subject classification*: primary 30F35, 11J70.

1. Introduction

Late in the 19th century, Markoff [11, 12] initiated an extensive theory of the minima of indefinite binary quadratic forms. He showed in particular that these begin with a countable discrete spectrum. Early in the 20th century, Ford [7] showed that these values are closely related to the geometry of the modular surface.

Some forty years later, Cohn [1] recognized a connection between these initial values of Markoff's spectrum and certain closed geodesics on the so-called homology cover of the modular surface. This was clarified and extended by various authors [8–10, 17, 18]. In particular, the Markoff numbers, which determine this initial countable set of values of the spectrum, correspond one-to-one to the simple closed geodesics on the hyperbolic once-punctured torus $\Gamma'\backslash\mathscr{H}$. The same result was shown to hold if $\Gamma'\backslash\mathscr{H}$ is replaced by $\Gamma(3)\backslash\mathscr{H}$ or $\Gamma^3\backslash\mathscr{H}$, the subject of our study.

Our results describe paths between elliptic fixed points of order 2. In addition, our work is global and metrical—indicating where on $\Gamma^3\backslash\mathscr{H}$ the paths run, their lengths,

and the order in which their constituent arcs are traversed. In particular, the fact that $\Gamma^3 \backslash \mathscr{H}$ has torsion singularities serves to put our work at some remove from the rather successful alternative approach for smooth surfaces—the laminations and train tracks of the Thurston school. See for instance, [3] and [16] for this. It should be noted that there are results using laminations for the behavior of geodesics near cusps (again for smooth surfaces), see say [13].

1.1. The geometry of $\Gamma^3 \backslash \mathscr{H}$ The modular group is $\Gamma = \mathrm{PSL}(2, \mathbb{Z})$. As a subgroup of $\mathrm{PSL}(2, \mathbb{R})$, Γ acts as möbius transformations on the Poincaré upper half-plane, $\mathscr{H}$. The modular surface is $\Gamma \backslash \mathscr{H}$. Let Γ' be the commutator subgroup of Γ. Then $\Gamma' \backslash \mathscr{H}$ is the punctured torus which plays the crucial role mentioned above.

Due to its simpler geometry, we focus upon a different cover of the modular surface. Let Γ^3 denote the subgroup of Γ generated by its cubes. This group has signature $(0; 2, 2, 2; \infty)$, that is, the corresponding surface is of genus zero with three elliptic fixed points of order two and a single puncture. It is traditional to refer to the puncture as the *cusp*; we refer to the elliptic fixed points as *efp2* or simply as *efp*. The group Γ^3 contains Γ' as an index two subgroup (for a discussion of this, see, for example, [18]). The fact that the surface is a sphere, and thus that one can easily apply notions related to the Jordan Curve Theorem, more than compensates for the complications due to ramification. Indeed, $\Gamma^3 \backslash \mathscr{H}$ has manageable geodesic geometry as the ramification is of degree 2; as is well known (see say [18]): the key feature is that a geodesic encountering an elliptic fixed point of order two 'bounces back' along the path of the encounter. For example, it happens that each simple closed geodesic on $\Gamma^3 \backslash \mathscr{H}$ connects a pair of distinct elliptic fixed points of order two.

The group of isometries of $\Gamma^3 \backslash \mathscr{H}$ is of order six. The three orientation preserving isometries are induced by the fundamental translation of Γ, that is, by the action on $\mathscr{H}$, of $S : z \mapsto z + 1$. To obtain the full group of isometries, one includes the orientation reversing map $z \mapsto -\bar{z}$.

The element of Γ, $T : z \mapsto -1/z$ is in Γ^3. Indeed, Γ^3 is generated by T_j with $j \in \{0, 1, 2\}$ where $T_j := S^j T S^{-j}$. Note also that $S^3 = T_2 T_1 T_0$ is in Γ^3. In particular, this implies that a geodesic on $\Gamma^3 \backslash \mathscr{H}$ of height greater than three must have a self-intersection. Here *height* means the greatest (or limsup) diameter of the set of lifts to $\mathscr{H}$ of the geodesic.

It is well known that neither one nor two geodesic arcs can bound a disc of trivial topology on a Riemann surface. We will say that an *illegal disc* is formed if a purported geodesic must be such that it would contradict this. On $\Gamma^3 \backslash \mathscr{H}$, a disc bounded by one geodesic arc (a 'monogon') or by two such arcs (a 'bigon') must contain at least the puncture or some elliptic fixed point. But, see [18], if there is exactly one elliptic fixed point and no cusp within a monogon, then the geodesic arc collapses onto the elliptic fixed point so as to have this point as a terminus.

Also, it follows easily (see [18]) that a simple closed geodesic on $\Gamma^3\backslash\mathcal{H}$, which perforce cuts the surface into two discs one of which contains exactly two efp2, must collapse onto an arc connecting these two efp2.

1.1.1. Automorphisms and the PSSI   The topological structure of $\Gamma'\backslash\mathcal{H}$ is that of a punctured torus. Thus, the fundamental group of $\Gamma'\backslash\mathcal{H}$ is isomorphic to the free group on two generators. But, Γ' has no elliptic elements and thus the fundamental group of $\Gamma'\backslash\mathcal{H}$ is isomorphic to Γ'. We can choose generators A and B for Γ' such that for the T_j defined above, $A = T_1 T_2 T_1 T_0$ and $B = T_0 T_1$. Indeed, these are the generators with which [4] work.

Geodesics correspond to conjugacy classes, closed geodesics to hyperbolic (that is, absolute value of trace greater than 2) classes. Hence, the fundamental group modulo inner automorphisms identifies the geodesics. All outer automorphisms, modulo inner automorphisms, of this fundamental group were shown by Nielsen [15] to be induced by homeomorphisms. Therefore, when proving a topological property holds throughout some automorphism group orbit of geodesics, it suffices to show that a single geodesic of the orbit enjoys the property.

Since geodesics realize the minimal number of self-intersections within their free homotopy class, it is easily seen that the simple closed geodesics on $\Gamma'\backslash\mathcal{H}$ form a single orbit under the action of the automorphism group. A PSSI is defined to be a closed geodesic of $\Gamma'\backslash\mathcal{H}$ which intersects itself exactly once and does not contain a monogon about the puncture. Crisp and Moran showed that these also form a single class modulo the action of the automorphism group. Indeed the class of the PSSI can be represented by $A^2 B^2$.

The orientation preserving outer automorphism group acts on the Teichmüller space of the hyperbolic once punctured tori as the Teichmüller group—basically, the action is given by a relabeling of geodesics with no other change in the underlying geometry. Furthermore, $\Gamma'\backslash\mathcal{H}$ geodesics in an orbit of the automorphism group of Γ' project to homeomorphic geodesics on $\Gamma^3\backslash\mathcal{H}$. We can simplify various topological arguments by simply checking that a property holds for a single representative of a geodesic configuration and thereby conclude that this topological property holds for every corresponding configuration under the action of the automorphism group. See, for instance, our proof of Lemma 1.

2. Results in this paper

NOTE. We use the abbreviation *SCG* for simple closed geodesic and continue with *efp* for elliptic fixed point.

2.1. Alignment and the Markoff equation   In Section 3 and Section 4, we show that there is a natural pairing between PSSI and SCG. Each PSSI has a unique simple

closed geodesic which it does not intersect, its paired SCG. Appropriately adding two more geodesics, we may triangulate $\Gamma^3 \backslash \mathcal{H}$. This set of curves can be lifted to $\mathcal{H}$, affording a fundamental region for Γ^3, one for each PSSI-SCG pair. In a precise sense, most of the area of $\Gamma^3 \backslash \mathcal{H}$ lies in the PSSI disc containing the cusp. Finally, the heights of the PSSI and the SCG are aligned with respect to the cusp. That is, they lie on a common vertical geodesic terminating in the cusp; highest lifts to the upper half-plane of the paired PSSI and SCG are concentric Euclidean semi-circles.

2.2. Two parametrizations of SCG on $\Gamma^3 \backslash \mathcal{H}$

2.2.1. Tine parametrization   In Section 5 we prove that a simple closed geodesic σ is characterized by its structure in neighbourhoods of the three elliptic fixed points. At each of the two elliptic fixed points which lie on σ, the structure is an even number of arcs (tines) around the elliptic fixed point and a single tine going into the elliptic fixed point; at the remaining elliptic fixed point there is a choice of passing above or below. Thus σ is be characterized by a signature $(m, n, \pm1)$. It turns out that $\gcd(2m + 1, 2n + 1) = 1$. There are $m + n$ arcs under or over the elliptic fixed point not on σ.

By using $\Gamma^3 \backslash \mathcal{H}$ isometries, we can normalize the signature to choose a representative of each isometry class of simple closed geodesics. We show how to pass between geodesics and normalized signatures, by way of recovering an element whose axis projects to the given simple closed geodesic. This is discussed in Section 5. From this, a minor refinement in the classic work of Harvey Cohn achieves a listing of all simple closed geodesics whose length lies between M and N, without simply determining all shorter geodesics (Section 5.3).

2.2.2. Basic arc parametrization   In Section 6 we describe a dynamic way of following the path of a SCG. This leads us to a parameterization of all isometry classes of SCG, and a signature identifying each such class. The signature is of the form (α, β, m, n) where α and β are chosen from our three basic SCGs (as discussed in Section 6), m is the number of laminations about α, and n is the number of times we choose to exit rather than terminate (when we have that option). This parametrization gives a good global picture of the SCG and also a simple upper bound for its length in terms of $l(\alpha), l(\beta), m$ and n. Here $l(\tau)$ is the length of the closed geodesic τ on $\Gamma^3 \backslash \mathcal{H}$.

3. Alignment of PSSI and SCG lifts

Choose an SCG τ connecting e_0 and e_1, say. There is then a unique geodesic connecting e_2 with itself that avoids τ and has its only self-intersection at e_2. This is straightforward, see our proof of Lemma 1. It turns out that these geodesics are

nothing but the proper single self-intersection *PSSI* geodesics of Crisp and Moran ([4]), projected to $\Gamma^3\backslash\mathcal{H}$.

Alternatively, starting with a PSSI, the corresponding SCG is the unique simple closed geodesic which does not intersect the PSSI. We call this the *parameter SCG* or *pSCG*.

Define a *generating pair* to be a pair of simple closed geodesics with a single point of intersection (the lifts of such a pair are the axes of a generating pair of elements in the free group Γ'). Both PSSI and parameter simple closed geodesics admit well defined highest points. Given such a PSSI-simple pair, we define its *marking segment* to be the shortest geodesic segment connecting their high points. As the following lemma shows, this allows us to associate to each PSSI a unique generating pair, which we will refer to as the *companions* of the PSSI.

LEMMA 1. *For each PSSI on $\Gamma^3\backslash\mathcal{H}$, there is a unique associated generating pair which intersects the PSSI and its parameter simple closed geodesic only at the elliptic fixed points while avoiding the marking segment.*

PROOF. Topologically, we can replace the marking segment by any segment joining the PSSI and its parameter simple closed geodesic (but avoiding the elliptic fixed points). Hence, by the automorphism group arguments, it suffices to consider the setting of the PSSI given by $T_3 T_0$. Here, the uniqueness of such a pair is geometrically clear; see Figure 1, where the pair is given by $T_0 T_1$ and $T_2 T_3$.   □

LEMMA 2. *Each PSSI of $\Gamma^3\backslash\mathcal{H}$ admits a highest lift on $\mathcal{H}$ which connects a pre-image of an elliptic fixed point of $\Gamma^3\backslash\mathcal{H}$ to the translate of this pre-image by $S^3 : z \mapsto z + 3$.*

PROOF. We simply note that a PSSI of $\Gamma^3\backslash\mathcal{H}$ forms a monogon about the puncture. Since the isotropy group in Γ^3 of ∞ is generated by S^3, the intersection point of this monogon lifts to some z and $z + 3$. But, that intersection point is an elliptic fixed point.   □

THEOREM 1. *If α is a PSSI on $\Gamma^3\backslash\mathcal{H}$, then there is an elliptic element $E \in \Gamma^3$ such that the axis of $S^3 E S^{-3} E$ is a highest lift of α. Furthermore, the axis of $S^3 E$ is a highest lift of the parameter simple closed geodesic.*

PROOF. Given α, by the previous lemma, there is a highest lift passing through the fixed point of an elliptic element of Γ^3, say that of E, and through the translation of this point by 3. This second point is fixed by $H = S^3 E S^{-3}$. Thus, $H E$ has as its axis the required lift.
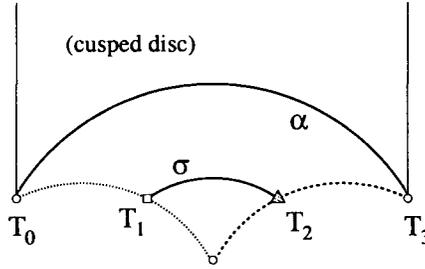
FIGURE 1. Axes of $T_3 T_0$ and $T_2 T_1$; a PSSI α and parameter simple σ

The parameter simple of α, say σ, has a highest lift which lies strictly between the vertical lines passing through the fixed points of E and H. Furthermore, the marking segment lifts to some geodesic segment joining the lifts of α and σ. This marking segment lift can be taken to lie above the lift of σ. But, this means that we can also lift the companions to the region. One companion thus lifts to join the fixed point of E to the lift of σ; the other lifts so as to join σ to the fixed point of H. These companion lifts meet the lift of σ at elliptic fixed points. Label these points as the fixed points of F and G respectively. We thus have identified a pentagon of vertices fixed by: S^3 (which fixes ∞), E, F, G, and H.

From the topology of the quadrilateral on the punctured sphere $\Gamma^3 \backslash \mathcal{H}$ bordered by the PSSI, parameter simple and the companions, a fundamental region for this surface is given by taking the pentagon and adjoining a triangle bounded by lifts of the companions and the parameter simple closed geodesic. That is, we can extend each of the companion lifts by a contiguous lift such that these two new lifts intersect, see Figure 1 for a special case. But, these contiguous lifts have new endpoints fixed by FEF and GHG. (To see this for the first of these, note that $FE = FEF \cdot F$.) Hence, $FEF = GHG$. Let $I = FEF$. The fundamental region is thus bounded by the fixed points of S^3, E, I and H. The elliptic F acts so as to glue the side of vertices fixed by E and I to itself; G acts similarly with respect to I and H. But this implies that $GFE = S^3$. Therefore, a highest lift of σ is indeed given by $S^3 E = GF$.   □

The largest hyperbolic punctured disc centered about the cusp of $\Gamma^3 \backslash \mathcal{H}$ is called the fundamental cuspidal horocycle. There is a natural striation of this: lift the fundamental horocycle to the upper half-plane with the cusp at ∞ and consider the striations induced by the vertical h-lines. Since the cusp is fixed only by translations, this induces a well-defined striation of the fundamental cuspidal horocycle. We say that points on $\Gamma^3 \backslash \mathcal{H}$ are *aligned* if they lie upon a single striation.

COROLLARY 1. *The high points of a PSSI and its parameter simple closed geodesic are aligned.*

PROOF. Given the PSSI, as above we can find an elliptic E. But, there exist integers m, k and b such that $E = \left(\begin{smallmatrix} -k & b \\ m & k \end{smallmatrix}\right)$. Thus, $S^3 E = \left(\begin{smallmatrix} 3m-k & 3k+b \\ m & k \end{smallmatrix}\right)$ and

$$HE = S^3 E S^{-3} E = \begin{pmatrix} 9m^2 - 3km + 1 & * \\ 3m^2 & 3km + 1 \end{pmatrix}.$$

Therefore, the axis of $S^3 E$ has endpoints $3/2 - k/m \pm \sqrt{9/4 - 1/m^2}$, while that of HE has endpoints $3/2 - k/m \pm \sqrt{9/4 + 1/m^2}$. The high points thus lift to lie on the line $x = 3/2 - k/m$.   □

4. PSSI-indexed fundamental regions for Γ^3

To begin, we introduce some geometric notation. Fix a PSSI and a highest lift of it; label the highest (left most) lift of its elliptic fixed point (that is, its point of self-intersection) as e_0. We label the endpoints of the (aligned, highest) lift of the pSCG as e_1 and e_2. Let $e' = T_{e_1}(e_0)$. (Note that here we are using the notation e_j in a more liberal fashion than previously.)

The boundary of our fundamental region consists of four h-arcs: one connecting e_0 to ∞; its translate by S^3, and two arcs connecting e_0 (respectively, $e_0 + 3$) with e'. Now the bisector of these last arcs are e_1 (respectively, e_2). Of course, e' is also the image of $e_0 + 3$ under T_{e_2}.

The first interesting feature of these fundamental regions is that they are parametrized by the PSSI-pSCG pairs. Since these are in 1-1 correspondence with the Teichmüller orbit of our surface, we have explicit lifts corresponding to this orbit. We know of no other nontrivial arithmetic surface for which such explicit fundamental regions have been given.

Secondly, quite naturally in terms of m, most of $\Gamma^3\backslash\mathscr{H}$ lies above the PSSI. Recall that SCG are in one-to-one correspondence with the Markoff triples (m, m_1, m_2), solutions in natural numbers to the Markoff equation $m^2 + m_1^2 + m_2^2 = 3mm_1m_2$. We say that a PSSI is associated with such a triple if its pSCG corresponds to the triple.

LEMMA 3. *The difference between the total area of $\Gamma^3\backslash\mathscr{H}$ and the region lying above the PSSI cusped disc associated with a Markoff triple (m, m_1, m_2) is $o(1/m)$.*

PROOF. We have that e_0 has height $1/m$ and the region above the PSSI lifts to a triangle with vertices at ∞, e_0 and $e_0 + 3$. The area of this triangle (and thus the region) is $\pi - 2\arctan 2/3m$. Now, the area of all of $\Gamma^3\backslash\mathscr{H}$ is π, three times that of the full modular surface, and we are done.   □

In closing this section, we note that this lemma suggests an attack on the notorious *Uniqueness Conjecture*—which states that m determines m_1 and m_2, apart from

trivialities—as follows: The heights on the PSSI and the pSCG are given solely in terms of m. If this occurred twice in the spectrum, there would be two accompanying fundamental regions given by the respective PSSI-pSCG pairs.

Consider the SL(2, ℝ) translation (meant literally: a parabolic action fixing ∞) placing one PSSI upon the other (isometrically). The previous lemma states that this is a surface isometry on all but a residue of the surface. Further, since the pSCG lifts are mapped one onto the other, the high points of the respective lifts are matched by this translation.

If the endpoint of the pSCGs were also matched by the translation, then translation would be an isometry on the full surface, and the Uniqueness Conjecture follows easily from this. However, though we know that the lengths of the pSCGs are the same, we do not know that the endpoints are the same. Since this issue is external to the PSSI cusped discs, which do match exactly, the fate of uniqueness seems in a way to reside on vanishingly small pieces of $\Gamma^3\backslash\mathscr{H}$.

Apropos of this, let us take this view: Say we specify e_0 and e_1. Then e' and more importantly e_2 are determined. Consider the conditions satisfied by e_2:

• It is an efp2 of Γ^3.
• It is the bisector of the arc between e' and $e_0 + 3$.
• The arc (e_1, e_2) is concentric with $(e_0, e_0 + 3)$.

Is it plausible that there are two different ways to do this? It appears to be quite difficult!

5. Paths and words of simple closed geodesics

5.1. Tine signatures of simple closed geodesics   We now associate an identifying *signature* to each isometry class of simple closed geodesics on $\Gamma^3\backslash\mathscr{H}$. This signature directly reflects the geometry of a normalized representative of the isometry class, with particular emphasis of the local geometry near the elliptic fixed points.

For ease of notation, we again label the three elliptic fixed points as e_0, e_1 and e_2. Given a simple closed geodesic, we may assume that it connects e_0 and e_2. The geodesic remains a bounded distance from the cusp, hence we can decompose $\Gamma^3\backslash\mathscr{H}$ into neighbourhoods of the cusp and of the e_i such that the simple geodesic is contained in the union of the three elliptic neighbourhoods. The intersection of the geodesic with each of the neighbourhoods of the e_i is a union of non-intersecting simple arcs. We call these e_i-arcs. We can decompose $\Gamma^3\backslash\mathscr{H}$ so as to ensure that the e_0-arcs do not meet e_2-arcs. Furthermore, we can choose our decomposition so as to minimize the number of e_0- and e_2-arcs.

We can envision the e_0-arcs and e_2-arcs as forming the *tines* of a pitchfork; the central tine meets the elliptic fixed point. Thus, there is an odd number of tines at
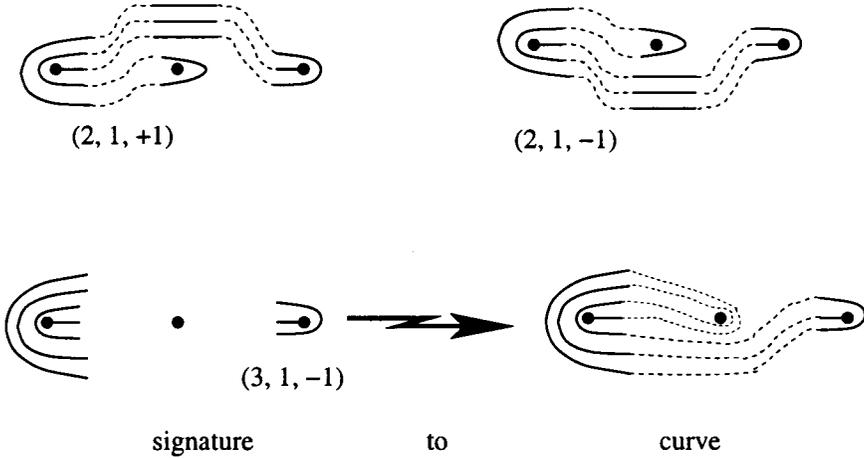
$(2, 1, +1)$ $(2, 1, -1)$

$(3, 1, -1)$

signature   to   curve

FIGURE 2. (m, n, ε) examples

each, say $2m + 1$ at e_0 and $2n + 1$ at e_2. By isometry, we may assume that $m \geq n$. We can also order each of the sets of tines and of arcs from bottom to top. Furthermore, we can note if an e_1-arc lies above or below e_1.

Each e_2-tine is joined to some e_1-arc. The remaining vertex of each of these e_1-arcs is joined to some e_0-tine. Thus, there remain $2(m - n)$ e_0-tines which must be joined two at a time by e_1-arcs. The loops so formed are homotopically interesting only if the e_0-tines joined by a single e_1-arc lie on opposite sides of e_0, one above, the other below. But, then the e_1-arcs meeting e_2-tines either all lie above e_1, or all lie below e_1. Our signature is $(m, n, +1)$ in this first case, $(m, n, -1)$ in the second.

Figure 2 gives some samples.

We note the following.

LEMMA 4. *Each isometry class of $\Gamma^3\backslash\mathscr{H}$ simple closed geodesic defines a unique signature.*

PROOF. The isometries of $\Gamma^3\backslash\mathscr{H}$ permute the e_j. Indeed, it is easily checked that $z \mapsto z + 1$ induces $\sigma(e_0, e_1, e_2) = (e_1, e_2, e_0)$ and $z \mapsto -\bar{z}$ induces $\tau(e_0, e_1, e_2) = (e_0, e_2, e_1)$. Note that this extends to a faithful representation of the isometry group as the full symmetric group on three letters.

Given a simple closed geodesic on $\Gamma^3\backslash\mathscr{H}$, this geodesic joins two of the e_j. There is a unique power of σ which takes these two to the pair e_0 and e_2. Furthermore, there is a unique element interchanging e_0 and e_2 (as a permutation, this is simply $\sigma\tau\sigma^{-1}$).

Hence, the normalization to a geodesic joining e_0 and e_2 such that there are more e_0-arcs than e_2-arcs uniquely identifies an isometry class representative of the class of

the initial simple closed geodesic. (If there are exactly as many e_0-arcs as e_2-arcs, it turns out that the isometry fixes the possible geodesics.)   □

We now show that the assignment of signature to simple closed geodesic isometry class gives a bijective correspondence.

LEMMA 5. *There is at most one isometry class of* $\Gamma^3\backslash\mathcal{H}$ *simple closed geodesic of any given signature.*

PROOF. Given a simple closed geodesic on $\Gamma^3\backslash\mathcal{H}$, we obtain the corresponding signature. Since a curve uniquely identifies its free homotopy class, and this latter uniquely identifies its isometry class, it suffices to show that there is only one way of joining the various e_i-arcs so as to obtain a simple, closed, connected curve from e_0 to e_2.

For each i, order the e_i-arc vertices by increasing height. To preserve simplicity, the joining of e_2-tines to e_1-arc vertices must respect this order. Similarly for e_0-tines. The only choice to be made is in the placement of the e_0-tines which are not joined by e_1-arcs to e_2-tines. Again, simplicity requires that these e_0-tines be adjacent in the order. But, were these not all at one of the extremes of this order, then the curve formed would either not be connected, or else m would not be minimal (in this latter case, we could homotope the curve so as to collapse e_0 tines). Now, the e_1-arcs joining e_0-tines must pass around e_1, else we can homotope the curve and show that m is not minimal. The sign of the signature indicates whether the other e_0-tines lie above or below these. Therefore, there is only one way of passing from a simple closed geodesic signature to a curve identifying an isometry class of such geodesics.   □

We now show exactly which signatures can be achieved by simple closed geodesics on $\Gamma^3\backslash\mathcal{H}$. All other signatures actually give laminations by simple closed curves.

THEOREM 2. *Each* $(m, n, \varepsilon) \in \mathbb{N} \times \mathbb{N} \times \{\pm 1\}$ *with* $m \geq n$ *defines a union of closed curves on* $\Gamma^3\backslash\mathcal{H}$. *A signature* (m, n, ε) *defines a simple closed curve if and only if* $\gcd(2m + 1, 2n + 1) = 1$.

PROOF. Given a signature, (m, n, ε), we draw $2m + 1$ tines at e_0, $2n + 1$ at e_2. We join the e_2-tines to e_0-tines by way of e_1-arcs. If $\varepsilon = 1$, then we join by starting from the top and working down. If $\varepsilon = -1$, we reverse this. All of these joining e_1-arcs lie either above or below e_1, again depending upon the sign of the signature.

We now join the remaining $2(m - n)$ e_0-tines to themselves by e_1-arcs which loop around e_1. This joining is done recursively, first the extreme (with respect to the order) tines are joined, then their neighbours, and so forth.

The above is always possible for any signature, and it is exactly this which recovers the simple curves. Now, suppose that there is more than one connected component of

the curve so formed. By a parity argument, the component meeting e_0 also meets e_2. Consider any other component. This new component must be a simple closed loop. But, ignoring any other components, this component is homotopically equivalent to the e_0—e_2 component. Hence, it has the same number of each of the e_i-arcs as that component. Of course, this is true for all of the components. But, this then implies that $\gcd(2m + 1, 2n + 1) > 1$.

If $\gcd(2m + 1, 2n + 1) =: d > 1$, then consider the curve of signature $((2m + 1)/d, (2n + 1)/d, \varepsilon)$. By the above, this is a simple closed curve. Placing $d - 1$ simple closed loops concentrically about this, each with the same e_i-arc data gives a union of curves of the signature (m, n, ε).   □

We call a signature (m, n, ε) a *simple* signature if $\gcd(2m + 1, 2n + 1) = 1$.

5.2. From signature to word   Given any simple signature (m, n, ε), we construct a matrix $W = W(m, n, \varepsilon)$ whose axis projects to a simple closed geodesic of this signature. The matrix W is given as a word in non-negative powers of the elements $a = T_0 T_1$, $b = T_0 T_2$ and $c = T_0 T_1 T_2 T_1$. Only two letters will appear in each word. To motivate the formation of W, we have the following result.

LEMMA 6. *Let γ be the curve formed by the above process from a simple signature (m, n, ε). Then γ is homologous to the projection of the axis of $a^{2(m-n)} b^{2n+1}$ when $\varepsilon = +1$ and to that of $a^{2(m-n)} c^{2n+1}$ otherwise.*

PROOF. The axis of a is the shortest path from e_0 to e_1. The axes of b and c project to the shortest paths which connect e_0 to e_2 while passing above and below e_1, respectively. Thus, we can collapse the various strands of γ onto multiples of a, b and c as announced.   □

In each homology class, there is at most one simple geodesic class, see [8] or [19] for this. But, each of (a, b) and (a, c) is a generating pair, as defined above. Hence, [2] gives the exact simple word in this class. Indeed, let

$$f_j := \lfloor j(2n + 1)/[2(m - n)] \rfloor - \lfloor (j - 1)(2n + 1)/[2(m - n)] \rfloor.$$

Then we have W given by

$$W(m, n, \varepsilon) := \begin{cases} \prod_{j=1}^{2(m-n)} a\, b^{f_j}, & \text{if } \varepsilon = +1; \\ \prod_{j=1}^{2(m-n)} a\, c^{f_j}, & \text{otherwise.} \end{cases}$$

Since $W(m, n, \varepsilon)$ is in the correct homology class, is simple, and there exists only one such simple, our construction gives the following.

PROPOSITION 1. *Let (m, n, ε) be a simple signature. Then the axis of $W(m, n, \varepsilon)$ projects to a simple closed geodesic on $\Gamma^3\backslash\mathcal{H}$ which is of signature (m, n, ε).*

5.3. All isometry classes of predetermined length   In this section we show that a minor extension of the work of Cohn in [2] suffices to produce a list of all non-isometric SCGs whose lengths lie between M and $M + T$. The point here is to does this without recursively generating the entire Markoff tree below $M + T$.

Like our signatures, the step-words of Cohn give a listing of words for simple closed geodesics, in his case the geodesics naturally lie on $\Gamma'\backslash\mathscr{H}$. The word $M(u, v)$ is defined as are our W above, except that now the two generating elements present in a word are the above mentioned A and B, with A corresponding to u. Note that A and B are of equal trace.

THEOREM 3. *Given integers u and v, let $M(u, v)$ be the step-word as above. Let $\gamma(u, v)$ be the projection to $\Gamma^3\backslash\mathscr{H}$ of the axis of $M(u, v)$. Then each simple closed geodesic of $\Gamma^3\backslash\mathscr{H}$ is some $\gamma(u, v)$ with u and v relatively prime or else with $uv = 0$. The $\gamma(u, v)$ such that $(u, v) = (0, 1)$ or else such that u and v re relatively prime with $0 < u < v/2$ uniquely represent all isometry classes of simple closed geodesics of $\Gamma^3\backslash\mathscr{H}$.*

PROOF. That every simple closed geodesic is of the form stated is proved in [2].

By inspection, T_0 conjugates each of A and B to its own inverse. The action of $S : z \mapsto z + 1$ takes (A, B) to (B^{-1}, AB), and the action of $z \mapsto -\bar{z}$ interchanges A and B. Using these, it is easy to see that every isometry class can be represented by an $M(u, v)$ with $0 < u < v$, or with $u = 0$ and $v = 1$. However, $M(u, v)$ minimizes trace amongst all words in A and B with the same ordered exponent sum (u, v), [8, 19]. Hence, applying the isometries S and $z \mapsto -\bar{z}$ sends $M(u, v)$ to a word of minimal trace and exponent sums $v - u$ and v. But, this word is simple, that is the axis of the corresponding element of Γ^3 projects to a simple on $\Gamma^3\backslash\mathscr{H}$. Therefore, this is conjugate to $M(v - u, v)$. Hence, we may assume that $0 < u < v/2$, or else $u = 0$ and $v = 1$. $\square$

We now give bounds on the lengths of the corresponding geodesics. We will need the fact that for matrices X and Y in SL(2, $\mathbb{R}$), tr$XY = $ tr$X$tr$Y - trX^{-1}Y$. Let $t(u, v) = trM(u, v)$.

LEMMA 7. *If $v > 0$ and $0 < u < v/2$, then $t(u, v) < t(u, v + 1)$.*

PROOF. From the formation of $M(u, v)$, there exists a unique i such that $f_i(u, v + 1) \neq f_i(u, v)$; for this i, $f_i(u, v + 1) = f_i(u, v) + 1$. For all j, let $f_j = f_j(u, v)$. Thus,

$$t(u, v + 1) = \text{tr}\left(\prod_{j=1}^{i} AB^{f_j} \cdot B \cdot \prod_{j=i+1}^{u} AB^{f_j}\right).$$

TABLE 1. All geodesic lengths between 111 and 112, 311 and 312, and 511 and 512.

111–112		311–312		511–512	
(u, v)	length, l(u, v)	(u, v)	length, l(u, v)	(u, v)	length, l(u, v)
(15, 61)	111.81105	(65, 174)	311.92698	(67, 279)	511.99248
(5, 59)	111.69281	(41, 170)	311.90176	(31, 272)	511.94472
(21, 62)	111.66743	(73, 175)	311.89694	(89, 283)	511.93460
(29, 63)	111.64199	(81, 176)	311.89247	(78, 281)	511.93390
(16, 61)	111.45688	(5, 163)	311.87693	(118, 287)	511.91655
(11, 60)	111.37087	(36, 169)	311.82771	(134, 289)	511.90738
(6, 59)	111.31817	(31, 168)	311.76366	(26, 271)	511.89306
(1, 58)	111.26651	(58, 173)	311.73204	(57, 277)	511.84437
(23, 62)	111.18520	(26, 167)	311.70931	(95, 284)	511.79665
(17, 61)	111.10797	(47, 171)	311.69539	(16, 269)	511.78973
*	*	(74, 175)	311.66115	(73, 280)	511.75977
511–512	*cont'd*	(21, 166)	311.65729	(103, 285)	511.75047
(85, 282)	511.41094	(16, 165)	311.60561	(11, 268)	511.73807
(53, 276)	511.4088	(11, 164)	311.55395	(47, 275)	511.72704
(48, 275)	511.35322	(53, 172)	311.52125	(111, 286)	511.70444
(97, 284)	511.31442	(6, 163)	311.50229	(127, 288)	511.67582
(38, 273)	511.24721	(59, 173)	311.49092	(135, 289)	511.67181
(91, 283)	511.23712	(37, 169)	311.45773	(143, 290)	511.66781
(80, 281)	511.23608	(1, 162)	311.45063	(1, 266)	511.63474
(113, 286)	511.22228	(67, 174)	311.44481	(68, 279)	511.62250
(121, 287)	511.20917	(83, 176)	311.42133	(37, 273)	511.62183
(137, 289)	511.20068	(27, 167)	311.33481	(90, 283)	511.58586
(33, 272)	511.19546	(60, 173)	311.24981	(79, 281)	511.58499
(28, 271)	511.14378	(17, 165)	311.23098	(63, 278)	511.54844
(59, 277)	511.10441	(76, 175)	311.18978	(27, 271)	511.51842
(23, 270)	511.09212	(43, 170)	311.17235	(104, 285)	511.50936
(18, 269)	511.04046	(7, 163)	311.12765	(58, 277)	511.47439
(106, 285)	511.02712	(38, 169)	311.08775	(120, 287)	511.44496
		(61, 173)	311.00869	(17, 269)	511.4151

But, since trace is a conjugacy class invariant, the trace relation for products leads to $t(u, v + 1) = 3t(u, v) - \operatorname{tr} V$, with V of ordered exponent sum $(u, v - 1)$. The minimal trace property thus gives that $t(u, v + 1) \geq 3t(u, v) - t(u, v - 1)$. We thus use induction, and it suffices to show that $t(u, 2u + 1) \geq t(u, 2u)$.

Let $r(u) = t(u, 2u)$ and $s(u) = t(u, 2u + 1)$. That is, $r(u) = \operatorname{tr}(A B^2)^u$ and $s(u) = \operatorname{tr}(A B^2)^u B$. Using the trace relation, it is easily seen that each of $r(u)$ and $s(u)$ forms a recurrence sequence. Explicit solutions then give that $s(u) > r(u)$ for $u \geq 1$.   □

LEMMA 8. *If $v > 0$ and $0 < u < (v + 1)/2$, then $t(u, v) > t(u + 1, v)$.*

PROOF. From the isometries, $t(u, v) = t(v, u + v)$ and similarly for $t(u + 1, v)$. From this, arguments as for the previous lemma give the result. $\qquad\square$

THEOREM 4. *Let l be the length of some simple closed geodesic on $\Gamma^3 \backslash \mathcal{H}$ and suppose that u, v as above are such that $uv \neq 0$ and where $l = l(u, v)$ is the length of $\gamma(u, v)$. Then there is a linear bound on l, depending only on v:*

$$c_1 v > l \geq c_2 v,$$

where $c_1 < 4v \log\left[(1 + \sqrt{5})/2\right]$ and $c_2 \geq (v/2)l(1, 2)$.

PROOF. Using recurrence sequences, it is readily seen that $t(1, v) = 3F_{2v}$, where F_j is the j-th Fibonacci number. Now, let $l(u, v)$ be the length of the $\Gamma^3 \backslash \mathcal{H}$ geodesic corresponding to $M(u, v)$. Thus, if $t = t(u, v)$, then $l(u, v) = 2\log(t + \sqrt{t^2 - 4})/2 < 2 \log t$. In particular, $l(1, v) < 2 \log 3F_{2v-2} < 4v \log[(1 + \sqrt{5})/2]$. By the preceding lemmas, this last gives an upper bound on $l(u, v)$.

Recall that [14] showed that geodesic length induces a norm on homology (of any hyperbolic punctured torus). From this in particular one has that $l(u, v) + l(v - u, v) \geq l(v, 2v)$. But, $l(v, 2v) = vl(1, 2)$. Furthermore, $l(u, v) + l(v - u, v) = 2l(u, v)$ (from our isometries). Therefore, $l(u, v) \geq (v/2)l(1, 2)$. $\qquad\square$

It is now straightforward to give all geodesics of length within predetermined bounds.

As an example of this, it is quickly checked that there is exactly one isometry class of geodesics of length between 11 and 12, $l(1, 6) = 11.475\ldots$. Table 1 gives the ten isometry classes and lengths between 111 and 112; the 29 isometry classes and lengths between 311 and 312; and the 47 isometry classes and lengths between 511 and 512. (There are 85 classes of length between 1011 and 1012.)

6. The array of SCG via the basic SCG

In this section, we give a second type of signature, again identifying each isometry class of simple closed geodesic. The emphasis here is more dynamic then in the previous section. In particular, the present signature naturally gives an elementary, but apparently quite good, lower bound on the length of the SCG in the class. Our construction of the signature naturally classes the SCG into arrays, determined by parametrizing *basic* geodesics.

Up to isometry, the shortest SCG are: α the projection of the axis of $T_1 T_2$, trace $= 3$, $l(\alpha) = \log(7 + 3\sqrt{5})/2 = 1.9248473\ldots$; β that of $T_0 T_2$, trace $= 6$, $l(\beta) = \log(17 + 3\sqrt{32}) = 3.525494345\ldots$; and finally γ of $T_0 T_1 T_0 T_2$, trace $= 15$, $l(\gamma) =$

$\log(223 + 15\sqrt{221})/2 = 5.407151662\ldots$. We call any element of the isometry classes of any of these three geodesics the *basic SCG*. Each of α, β and γ has an element of its isometry class connecting any pair of distinct elliptic fixed points.

To create an array of SCG connecting e_0 and e_2, we choose a *hinged pair* of basic SCG. Such a pair shares a single point (e_1 or e_2). The geometry of our construction relies upon the old insight ([18]) that SCGs on $\Gamma^3\backslash\mathcal{H}$ exist only because it is possible to use the efp2s as 'maypoles' in order to reverse direction, without incurring an intersection. In a typical SCG, all three efp2s are used in this way, with two of them also being termini. We have two geometric steps—*entry/exit* over one of the basic SCG in the hinged pair and *lamination* on the other.

We now create an example array by first choosing a hinged pair μ, connecting e_0 and e_2, and λ, connecting e_1 and e_2. The m, n-entry of the array is the SCG formed in the following manner. First, upon leaving e_0 we approach e_2, adhering to the path of μ. We now wrap about λ a total of m times. We stop facing e_2. We can terminate or circle the point e_2 in the opposite direction, which then forces an exiting process comprised of m more wraps about λ. (We have now drawn the *labyrinth* of the SCG, see Figure 3.) Upon exiting the labyrinth, there is a unique direction around which we can circle e_0 one time (only!) before entering the labyrinth again in such as way as to have access to e_2 and e_1. This single laminate adheres to our chosen path from e_0 to e_2, say μ. We have only one further choice: the number of times to exit $(n-1)$ before choosing termination when facing e_2 or e_1.

Thus we define the *signature* of a SCG so constructed to be (λ, μ, m, n).

It is a simple matter to describe approximately where the height occurs on an SCG corresponding to an entry of the array: If there is an exit, then the height occurs after the last exit and before the last entry. Depending on m and n, this can lie anywhere on the geodesic. If there is no exit $(n = 1)$, the height occurs before we start spinning about λ.

THEOREM 5. *Fix a pair of basic SCG. The signatures* (λ, μ, m, n) *identify each SCG homotopy class once, up to isometry. Given a SCG of signature* (λ, μ, m, n), *its highest point occurs after the last exit and before the last entry, as defined above, except in those cases where there is no exit, when it occurs before the (unique) entry. The number* $n \, 1 \, \mu + [m + (n-1)(2m-1)] \, 1 \, \lambda$ *is an upper bound for the length of the geodesic of that position in the array.*

PROOF. We first show that each isometry class does indeed correspond to (at least) one of the signatures. The main step is of course to identify the associated hinged pair.

Suppose an SCG cut is given. By applying isometries, we may assume that it connects e_0 to e_2 and that the number of arcs at e_2 is greater than the number at e_0.
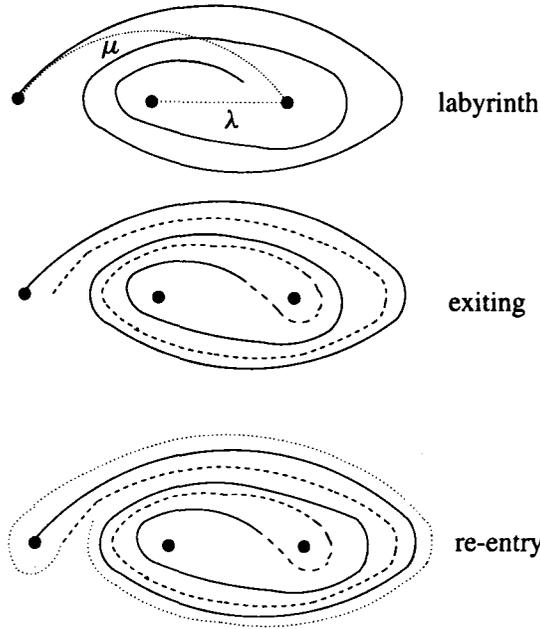
FIGURE 3. A labyrinth and its basic SCG

TABLE 2. Comparing naive upper bounds and true lengths: $\lambda = \alpha$; $\mu = \beta$. (Beginning of array)

upper bound				true length			
5.45034	7.37519	9.30004	11.2249	5.40715	7.32581	9.24975	11.1745
10.9007	16.6752	22.4498	28.2243	10.8143	16.5756	22.3489	28.1234
16.351	25.9753	35.5995	45.2237	16.2215	25.8253	35.4482	45.0724
21.8014	35.2753	48.7492	62.2232	21.6286	35.0751	48.5475	62.0214
27.2517	44.5753	61.899	79.2226	27.0358	44.3248	61.6468	78.9704

Follow the SCG away from e_0 until its first intersection with the vertical geodesic from e_2 to the cusp. (Vertical here means that this geodesic is the projection of a vertical h-line.) The homotopy class of the union of this arc and of the arc of the vertical geodesic joining it to e_2 is a simple class; let λ be the geodesic in this class.

Form the largest disk about e_1 which does not meet the SCG; mark one point where the closure of this disk meets the SCG. Cut the SCG at this point. Follow each of the two newly formed arcs away from e_1 until its closest point to e_2. Of these two arcs, take the arc whose closest point is nearer to e_2. Connect this nearer point to e_1; the union of this arc with the chosen arc gives the class of μ.

Note that λ and μ meet at e_2. Since the unions of arcs which lead to these geodesics did not intersect but at these points, neither do λ and μ have any other points of intersection. By this construction, λ and μ have no 'swings about the maypoles' of

TABLE 3. Comparing naive upper bounds and true lengths : $\lambda = \alpha$; $\mu = \beta$. ($23 \leq n, m \leq 25$)

upper bound			true length		
2030.95668	2117.57481	2204.192942	2029.79664	2116.41477	2203.03290
2121.10031	2211.56813	2302.03595	2119.88983	2210.357650	2300.82547
2211.24393	2305.56145	2399.87897	2209.98301	2304.30053	2398.61805

the efp2. However, it is easily checked that the only SCG of this simplicity are the basic SCG.

We must now show that the given SCG can be arise from our constructed using the hinged pair λ and μ. First note that it is possible that the SCG is simply λ; this is the case of signature (λ, μ, 1, 0). In all other cases, μ is indeed of geometric significance. For the geodesic to remain connected and simple, we easily find that it is formed by a process as in our construction.

We now arrange the SCG in an array, by indicating the associated homology class. The initial entry, the geodesic in the isometry class of μ which connects e_0 to e_2 and is the sole entry of the $n = 0$-column, we suppress.

$$
\begin{array}{ccccccccc}
\mu + \lambda & \rightarrow & \mu + 2\lambda & \rightarrow & \mu + 3\lambda & \rightarrow & \mu + 4\lambda & \rightarrow & \cdots \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
2\mu + 2\lambda & \rightarrow & 2\mu + 5\lambda & \rightarrow & 2\mu + 8\lambda & \rightarrow & 2\mu + 11\lambda & \rightarrow & \cdots \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
3\mu + 3\lambda & \rightarrow & 3\mu + 8\lambda & \rightarrow & 3\mu + 13\lambda & \rightarrow & 3\mu + 18\lambda & \rightarrow & \cdots \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
\cdots & \rightarrow & \cdots & \rightarrow & \cdots & \rightarrow & \cdots & \rightarrow & \cdots
\end{array}
$$

The entries in the array count the number of arcs in the free homotopy class of λ and μ which is required to make the corresponding SCG. The algorithm for generating this array is simple: In the n, m-th position, put $n\mu + [m + (n - 1)(2m - 1)]\lambda$. Here, to reiterate, m is the number of initial wraps about λ and $n - 1$ is the number of exits.

Note also that if we start at the n, m-th position and rise k rows ($0 < k < n$, of course), then the coefficient of μ decrements by k and that of λ, by $(2m - 1)k$. Likewise, a move to the right l columns leaves the coefficient of μ unchanged and while that of λ increments by $(2n - 1)l$.

Since a geodesic minimizes length in its homotopy class, it is clear that the geodesic of entry n, m has length at most $n\,l\,(\mu) + [m + (n - 1)(2m - 1)]\,l\,(\lambda)$. $\square$

It is perhaps surprising that this upper bound is in fact a fairly good approximation, as Table 2—where we give values of $1 \leq m, n \leq 5$—and Table 3 indicate. We leave a more detailed pursuit of the quality of this approximation to a later publication.

References

[1] H. Cohn, 'Approach to Markoff's minimal forms through modular functions', *Ann. of Math. (2)* **61** (1955), 1–12.

[2] ——, 'Markoff geodesics in matrix theory', in: *Number theory with an emphasis on the Markoff spectrum* (eds. A. D. Pollington and W. Moran), Lecture Notes in Pure and Appl. Math. 147 (Marcel Dekker, New York, 1993) pp. 69–82.

[3] A. J. Casson and S. A. Bleiler, *Automorphisms of surfaces after Nielsen and Thurston*, London Math. Soc. Student Texts 9 (Cambridge University Press, Cambridge, 1988).

[4] D. J. Crisp and W. Moran, 'Single self-intersection geodesics and the Markoff spectrum', in: *Number theory with an emphasis on the Markoff spectrum* (eds. A. D. Pollington and W. Moran), Lecture Notes in Pure and Appl. Math. 147 (Marcel Dekker, New York, 1993) pp. 83–93.

[5] ——, 'The Markoff spectrum and closed geodesics with one self-intersection on a torus', preprint, (Flinders University, 1995).

[6] T. W. Cusick and M. E. Flahive, *The Markoff and Lagrange spectra*, Mathematical Surveys and Monographs 30 (Amer. Math. Society, Providence R.I., 1989).

[7] L. E. Ford, 'A geometrical proof of a theorem of Hurwitz', *Proc. Edinburgh Math. Soc.* **35** (1917), 59–65.

[8] A. Haas, 'Diophantine approximation on hyperbolic Riemann surfaces', *Acta Math.* **156** (1986), 33–82.

[9] ——, 'Diophantine approximation on hyperbolic orbifolds', *Duke Math. J.* **56** (1988), 531–547.

[10] J. Lehner and M. Sheingorn, 'Simple closed geodesics on $H^+/\Gamma(3)$ arise from the Markov spectrum', *Bull. Amer. Math. Soc. (N.S.)* **11** (1984), 359–362.

[11] A. A. Markoff, 'Sur les formes quadratiques binaires indéfinie', *Math. Ann.* **15** (1879), 381–406.

[12] ——, 'Sur les formes quadratiques binaires indéfinies II', *Math. Ann.* **17** (1880), 379–400.

[13] G. McShane, 'Simple geodesics and a series constant over Teichmüller space', *Inventiones Math.* **132** (1998), 607–632.

[14] G. McShane and I. Rivin, 'A norm on homology of surfaces and counting simple geodesics', *Intl. Math. Res. Not.* **2** (1995), 61–69.

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

[16] R. C. Penner and J. L. Harer, *Combinatorics of train tracks*, Ann. of Math. Studies 125 (Princeton University Press, Princeton, N.J., 1992).

[17] C. Series, 'The geometry of Markoff numbers', *Math. Intelligencer* **7** (1985), 20–29.

[18] M. Sheingorn, 'Characterization of simple closed geodesics on Fricke surfaces', *Duke Math. J.* **52** (1985), 535–545.

[19] H. Zieschang, 'Minimal geodesics of a torus with a hole', *Izv. Akad. Nauk USSR* **50** (1986), 1097–1105; English translation: *Math. USSR Izv.* **29** (1987), 449–457.

Oregon State University   CUNY - Baruch College
Corvallis, OR 97331   New York, NY 10010
USA   USA
e-mail: toms@math.orst.edu   e-mail: marksh@alum.dartmouth.org