CLASSIFYING LOW HEIGHT GEODESICS ON $\Gamma^3\backslash\mathcal{H}$

THOMAS A. SCHMIDT AND MARK SHEINGORN

For Marvin Knopp, our teacher and loyal friend for $17 + 42$ years.
With "Thanks for the Memories."

Abstract. We show that low height-achieving non-simple geodesics on a low-index cover of the modular surface can be classified into seven types, according to the topology of highest arcs.

The lowest geodesics of the signature $(0; 2, 2, 2, \infty)$-orbifold $\Gamma^3\backslash\mathcal{H}$ are the simple closed geodesics; these are indexed up to isometry by Markoff triples of positive integers (x, y, z) with $x^2 + y^2 + z^2 = 3xyz$, and have heights $\sqrt{9 - 4/z^2}$. Geodesics considered by Crisp and Moran have heights $\sqrt{9 + 4/z^2}$; they conjectured that these heights, which lie in the "mysterious region" between 3 and the Hall ray, are isolated in the Markoff Spectrum. As a step in resolving this conjecture, we characterize the geometry on $\Gamma^3\backslash\mathcal{H}$ of geodesic arcs with heights strictly between 3 and 6. Of these, one type of geodesic arc cannot realize the height of any geodesic.

1. Introduction and Statement of Results

H. Cohn [Co] observed that the seemingly purely number theoretic Markoff spectrum formed by the minima on the integer lattice of (appropriately normalized) indefinite binary quadratic integral forms can be expressed in terms of heights of geodesics of certain small-index covers of the modular surface, $\Gamma\backslash\mathcal{H}$; see also [Se], [LS], [BLS], [H], [H2], as well as chapter seven of [CF]. Indeed, the initial segment of the Markoff spectrum is closely related to the heights of the simple closed geodesics of the hyperbolic punctured torus uniformized by the commutator subgroup of Γ. This relationship also holds with respect to the heights of the simple closed geodesics of $\Gamma^3\backslash\mathcal{H}$, the quotient of this torus by its elliptic involution [Sh].

The height of a geodesic reports its penetration into a cusp; normalizing so that this cusp lifts to ∞, the height of a geodesic is the supremum of the Euclidean diameters of lifts to $\mathcal{H}$ of the geodesic. The set of all heights of geodesics of $\Gamma^3\backslash\mathcal{H}$ gives the Markoff spectrum.

An approach to the geometric study of the Markoff spectrum has emerged: study the simple closed geodesics, thereafter the singly self-intersecting geodesics, and so on with increasing self-intersection number, [Se], [Cr],[CM], [CM2], [CrEtAl]. Here we ask: To what extent does the height of a geodesic determine topological aspects of the geodesic?

The well-known fact that a geodesic of $\Gamma^3\backslash\mathcal{H}$ of height at least $3n$, for a natural number n, must have at least n self-intersections already illustrates constrained height restricting topology.

A geodesic with a 'highest arc' is called a height-achieving geodesic, these are of main interest for the Markoff spectrum, see Lemma 1. We show that there are eight types of height achieving arcs in our region of interest, see Figure 4, on page 8. Furthermore, a highest arc can never be of the eighth of these types, see §5.2.

Date: 26 July, 2007.
2000 *Mathematics Subject Classification.* 30F35, 11J70.
Key words and phrases. Geodesics, Markoff spectrum, modular surface, geodesic configurations.

Main Theorem . *Let γ be a height-achieving geodesic on $\Gamma^3\backslash\mathcal{H}$ of height strictly between $h = 3$ and $h = 6$. Then γ has a highest arc of one of the first seven of the types specified in Definition 5.*

1.1. **Remark on Generality: Teichmüller Space Context.** The classification given by our Main Theorem applies mutatis mutandis to the quotient of any hyperbolic once-punctured torus by its elliptic involution, see [Sh] or [Sch] for the necessary background for this generalization.

1.2. **Outline.** Further background and motivation are given in § 2 and § 3, Section 4 gives the technical results used in § 5 to establish the division into eight types of high arcs, Theorem 1. In §5.2 we show that arcs of Type 8 can never be highest arcs.

1.3. **Related Work.** In [SS3], we use the Markoff triples-indexed fundamental domains for Γ^3 introduced in [SS2] to study the various types of low height-achieving geodesics. In particular, we show that our Type 3 geodesics are surprisingly constrained: each is of height $\sqrt{9/4 + 1/(z\,a_n\,)^2}$ with z a greatest value in a Markoff triple, and for each z, $a_n = a_n(z)$ an explicit increasing sequence of integers.

1.4. **Acknowledgements.** We thank D. Crisp and W. Moran as well as M. Flahive for their encouragement and interest in this and related work. We also thank the referee for a careful reading.

2. Markoff Spectrum and Lifts of Geodesics

Throughout, we use the term *geodesic* to refer to a complete infinite geodesic; a *geodesic ray* is a half-infinite ray; finally, a *geodesic arc* is a finite segment lying along some geodesic (which we assume to be closed unless otherwise stated). Recall that the geodesics of the Poincaré upper half-plane, $\mathcal{H}$, are the vertical (half-)lines and the semi-circles centered on the real line. Following tradition, we refer to these as *h-lines*, and their intersections with the (extended) real line as *feet*, with the left foot of a non-vertical h-line being less than its right foot. The highest point of a non-vertical h-line is called its *apex*.

Let $\mathcal{S}$ be a hyperbolic surface (or orbifold) with a distinguished cusp. One can uniformize $\mathcal{S}$ such that the cusp of $\mathcal{S}$ is given by ∞. A *high point* of a geodesic arc on $\mathcal{S}$ is a point on the arc that lifts to the apex of some h-line; in other words, the unit tangent vector to the geodesic based at this point can naturally be viewed as being horizontal. The *height* of the point is the Euclidean radius of this h-line. A *highest point* of a geodesic is a point along the geodesic that lifts to be as least as high as any other point on the geodesic. The *height* of a geodesic is the supremum of the Euclidean diameters of the h-lines that cover it. (Warning: a point $p \in \mathcal{H}$ has height equal to $\Im(p)$, the h-line of apex p has height $2\,\Im(p)$.)

Suppressing further reference to our particular cusp, we define the *Markoff spectrum* of $\mathcal{S}$ to be the set of all heights of geodesics of $\mathcal{S}$. The related notion of the *geometric Markoff value*, see [H3], of a geodesic γ is defined as the area of the largest horocycle centered at the cusp which is disjoint from γ. Up to a normalizing constant, the height of γ is the inverse of its geometric Markoff value.

We say that a geodesic γ *achieves its height* if there is a lift of γ that has diameter equal to the height of γ. Due to the following result, we restrict our attention the height achieving geodesics. Our lemma, similar to a result in [Cr], generalizes Lemma 6 of Chapter 1 of [CF].

Lemma 1. *Let $\mathcal{S}$ be a hyperbolic surface (or orbifold) with a distinguished cusp, and let γ be a geodesic of $\mathcal{S}$. If the height of γ is finite, then there is a height-achieving geodesic of $\mathcal{S}$ of the same height as γ.*

Proof: Of course, if γ is itself height-achieving, we are done. Otherwise, we normalize so that the cusp is represented by ∞, and suppose that the fundamental translation is $w \mapsto w + \lambda$. We thus fix a strip in $\mathcal{H}$ of Euclidean width λ, and need only consider lifts of γ in this strip. Let γ be of height h. Since h is the supremum of diameters of lifts of γ, there is a sequence of lifts of γ of Euclidean diameter converging from below to the height of γ; we may assume that the apexes of these lifts lie in some compact rectangle. A subsequence of these apexes converges to a point, say p, lying on the horocycle of equation $y = h/2$.

The limit point p is the apex of some h-line, ℓ. We claim that ℓ projects to a geodesic of $\mathcal{S}$ that has the same height as γ. It suffices to show that ℓ is a highest lift of this projection. If not, then there exist $\epsilon > 0$ and some element M of the Fuchsian group uniformizing $\mathcal{S}$ such that M sends ℓ to an h-line of apex lying above $y = h/2 + \epsilon$. But, the action of M is continuous on $\mathcal{H}$ — there is some δ such that M sends any h-line of apex within δ of p to an h-line of apex lying above $y = h/2 + \epsilon/2$. This is a contradiction, as h is the supremum of heights of lifts of γ. □

3. Geometry of the Surface

The modular group is $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$. It acts on the Poincaré upper half-plane $\mathcal{H}$ by way of möbius transformations; the *modular surface* is $\Gamma\backslash\mathcal{H}$. Let Γ' be the commutator subgroup of Γ. Then $\Gamma'\backslash\mathcal{H}$ is a punctured torus. Due to its simpler geometry, we focus upon a different cover of the modular surface — the quotient of $\Gamma'\backslash\mathcal{H}$ by its elliptic involution.

3.1. Group and Surface Basics.

The elliptic involution has three fixed points, the Weierstrass points. The quotient of this hyperbolic surface by this order two automorphism is a hyperbolic orbifold of genus zero with three singularities and a single puncture. This quotient is $\Gamma^3\backslash\mathcal{H}$ — where, as usual, Γ^3 denotes the subgroup of Γ generated by its cubes — the group Γ^3 thus has signature $(0; 2, 2, 2; \infty)$ — and contains Γ' as an index two subgroup (for a discussion of this, see, for example, [Sh]). The singularities are the projections of elliptic fixed points of order two for Γ^3, we often refer to them simply as *elliptic points*. Highest lifts to $\mathcal{H}$ of these singularities are i, $1 + i$ and $2 + i$, respectively. It is traditional to refer to the puncture as the *cusp*. We call all other points of $\Gamma^3\backslash\mathcal{H}$ *regular* points. Each simple closed geodesic on $\Gamma^3\backslash\mathcal{H}$ connects a pair of distinct elliptic points of order two, see [Sh] (especially, Figure 1.3b there) or [H2]; this follows from the fact that each simple closed geodesic of a hyperbolic punctured torus meets two distinct Weierstrass points.

The element of Γ, $T : w \mapsto -1/w$ is in Γ^3. Indeed, Γ^3 is generated by T_j with $j \in \{0, 1, 2\}$ where $T_j := S^j T S^{-j}$ with $S : w \mapsto w + 1$ the fundamental translation of Γ itself. The fixed points of the T_j project to the three elliptic points of $\Gamma^3\backslash\mathcal{H}$. Note also that $S^3 = T_2T_1T_0$ is in Γ^3. It is the presence of this translation in the group that ensures that any geodesic on $\Gamma^3\backslash\mathcal{H}$ of height greater than three has a self-intersection.

Throughout, we will tacitly invoke the Jordan curve theorem in our setting. We classify geodesics by geodesic arc configurations; since our surface is a once-punctured sphere with three marked points, these are easily envisioned.

3.2. Geodesic Configurations.

Definition 1. (a.) A *disk* on a hyperbolic orbifold or surface is a region isometric to a standard finite area hyperbolic disk. (In particular, a disk contains no singularities.)

(b.) A *monogon* on a hyperbolic orbifold or surface is a closed loop formed by a single (otherwise simple) geodesic arc.

(c.) A *bigon* is a closed loop formed by two simple geodesic arcs, meeting in at most two distinct points.

(d.) We say that a monogon, or a bigon, is *about the cusp* if one of the connected components of its complement contains a cusp and no elliptic points.
(e.) We say that a geodesic or configuration of geodesic arcs is *pointed* if it includes an elliptic point.
(f.) We refer to the points of (self-)intersection of a monogon or bigon as *nodes*. The node of a pointed monogon is its elliptic point (there is indeed at most one).
(g.) A geodesic including an unpointed monogon continues as a ray from the node, the oppositely oriented geodesic also continues as a ray from this node. We call these rays the *arms* extending the monogon. See Figure 3 on page 7.

The following result eliminates trivial cases in upcoming proofs.

Lemma 2. *Suppose that α is a monogon on some hyperbolic surface or orbifold, and that α bounds a cusped disk. Then the node of α is not a high point.*

Proof. A high point lifts to $\mathcal{H}$ to be the apex of an h-line; that is, it lifts to be the base point of a horizontal tangent vector. However, it is clear that the two unit tangent vectors of α at its node cannot both lift to be horizontal. That is, the path of α continues to a higher point than the node. □

The following is well-known. Items (a) and (b) result from considering lifts to $\mathcal{H}$. Item (c) is discussed in [Sh]; recall that a geodesic ray encountering an elliptic point of order two is reflected — the geodesic retraces its path, but in the opposite direction. For ease of discussion, we say that a geodesic ray has its *path terminating* at an elliptic point if the ray meets this point. For pictorial indications of applications of this lemma, see Figures 3 and 5.

Lemma 3. *Suppose that $\mathcal{S}$ is a hyperbolic orbifold or surface.*
(a.) *A monogon of $\mathcal{S}$ cannot bound a disk.*
(b.) *A bigon of $\mathcal{S}$ cannot bound a disk.*
(c.) *If in the homotopy class of a geodesic of $\mathcal{S}$ there is a curve including a loop about an elliptic point of order two, then the path of the geodesic terminates at that elliptic point.*

3.3. **Simple Rays on the Surface.** Haas, in Lemma 5.1 of [H], showed that on any hyperbolic surface with finitely many cusps, any simple ray that does not escape into a cusp has a lift to $\mathcal{H}$ that shares a foot with a lift of a geodesic in the closure of the simple closed geodesics. Elements of this closure are examples of Thurston's geodesic laminations, see [CEG] or [Bon]. In briefest terms, a lamination is a closed subset given as the disjoint union of simple geodesics, its *leaves*; a lamination is *minimal* if it is the closure of each of its leaves; following McShane [Mc], here a lamination is called *irrational* if it is a non-empty minimal compact lamination that is not a closed geodesic. If a geodesic ray shares a foot of a lift with some other geodesic ray, then it *spirals into* this second ray; if the second ray lies on a finite length geodesic, then the first ray comes arbitrarily close to every point of its path. Recall that $\Gamma^3\backslash\mathcal{H}$ is the quotient of the modular punctured torus by its elliptic involution; upon taking this quotient, the Haas result gives the following.

Lemma 4. *(Haas) If ρ is a simple geodesic ray on $\Gamma^3\backslash\mathcal{H}$, then exactly one of the following holds:*
(a.) *the path of ρ terminates at an elliptic point;*
(b.) *ρ escapes in the cusp;*
(c.) *ρ spirals into a simple closed geodesic;*
(d.) *the set of limit points of ρ is an irrational lamination.*

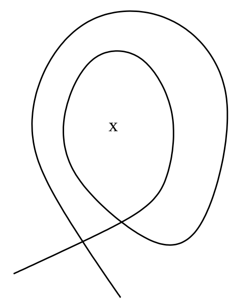

FIGURE 1. Double loops about the cusp are high.

Since any h-line of finite height greater than 3 projects to a $\Gamma^3\backslash\mathcal{H}$ geodesic with self-intersection, we have the following.

Corollary 1. *A geodesic of* $\Gamma^3\backslash\mathcal{H}$ *that does not escape into the cusp is simple if and only if its height is at most* 3.

By the Haas result, any infinite simple geodesic ray eventually comes arbitrarily close to the path of an infinite simple geodesic. But any such simple geodesic is the limit of simple closed geodesics, we can take a subsequence in which each simple closed geodesic joins the same two elliptic points of $\Gamma^3\backslash\mathcal{H}$. Therefore, the infinite simple geodesic, and hence the infinite simple ray, comes arbitrarily close to these two elliptic points. This gives the following corollary, which is related to Lemma 4.1 of [Mc].

Corollary 2. *If* ρ *is an infinite simple geodesic ray on* $\Gamma^3\backslash\mathcal{H}$ *, then* ρ *comes arbitrarily close to at least two elliptic points.*

For ease of reference, we define a specific type of arc.

Definition 2. A geodesic arc is a *rho* if it is either a monogon, or consists of a simple arc, called the *stem* of the rho followed by a monogon, with stem and monogon meeting only at the node of the monogon. (See the representation of a Type 3 arc in Figure 4 of page 8.)

Definition 3. We say that a geodesic on a hyperbolic surface or orbifold *makes a double loop about a cusp* if it contains an arc α that is a rho, whose monogon is about the cusp, and is continued by a simple arc that meets the stem of the rho so as to form an unpointed bigon about the cusp. See Figure 1.

A geodesic that makes a double loop about a cusp is covered by an h-line that intersects its own image under twice the fundamental translation corresponding to the cusp. This gives the following.

Lemma 5. *A geodesic of* $\Gamma^3\backslash\mathcal{H}$ *that makes a double loop about the cusp has height at least 6.*

4. EXTENDING BY RAYS

This is perhaps our most technical section. In the next section, we require it in the proof of our main classification result, Theorem 1, where we decompose the path of a geodesic passing through a point into two rays. Here we examine how geodesic rays emanating from a single point can meet. We show that any two rays emanating from a single point admit subarcs such that the union is one of only a few possible basic configurations.

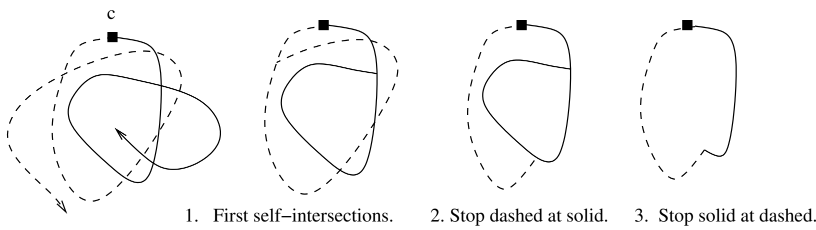

FIGURE 2. Finite configuration from rays.

The basic configurations are found by trimming rays at self-intersections, and then again at intersections with each other. See Figure 2 for a sketch of an application of this process. Note that here and elsewhere, there is an obvious difficulty: an infinite ray ρ could intersect a given arc α so that there is no first point of intersection as measured along α. However, as α is of finite length, there is a first point of intersection as measured along ρ.

Note that in the statement and proof of this classification, *arc* may mean infinite geodesic segment.

Lemma 6. *Suppose that c is a regular point of $\Gamma^3\backslash\mathcal{H}$ and that ρ_1 and ρ_2 are distinct geodesic rays emanating from c that do not escape into the cusp. If the union of the rays is non-simple, then there are connected subarcs emanating from c, α_1 and α_2, of ρ_1 and ρ_2 respectively, such that both*
(a) at least one of the following holds:

(1.) *each α_i is a rho;*
(2.) *exactly one α_i is a rho, the other is simple and joins c to an elliptic point;*
(3.) *the union of the α_i forms a simple loop;*
(4.) *exactly one of the α_i is a rho, the other is an infinite simple ray;*

and,
(b) other than in case (3), the α_i meet only at c.

Proof. If either ρ_i is simple with path terminating at an elliptic point, accordingly let α_i denote this finite length arc. If either ray is non-simple, then it contains a shortest geodesic arc that begins at c and has exactly one point of self-intersection; accordingly let α_i be this arc.

Unless both of the original rays is infinite simple, we have determined at least one finite length arc; we may assume that this is α_1. If ρ_2 meets α_1, we now let α_2 be the shortest subarc of ρ_2 emanating from c that meets α_1; thereafter, replace α_1 by its shortest subarc emanating from c that meets α_2. The arcs so formed cannot terminate at distinct elliptic points, for then the rays ρ_i would have already done so, and would have a simple union. If both arcs terminate at the same elliptic point, we have an instance of case (3). Thus, we assume in what follows that at least one arc does not end at an elliptic point. We now have one of the following: each arc meets itself, the two arcs meet only at c, this is an instance of case (1); one arc terminates in an elliptic point, the other arc meets itself, the two arcs meet only at c, of case (2); or, the arcs meet (before either meets itself), an instance of case (3). If some ray remains infinite after the above process, then it must not be met by the other subarc. Since the union is non-simple, it must be that this second arc meets itself, thus the configuration is of case (4).

Finally, if both rays are infinite simple, then since their union is non-simple, the rays meet. Choose a finite length subarc of ρ_1 emanating from c that includes intersection points with ρ_2;

then ρ_2 has a first point of intersection with this finite arc. Let each α_i be the subarc of ρ_i running from c to this point, this is of case (3). □

With a minor abuse of terminology, we refer to any element of the collection of elliptic points and the cusp as a *special point.* By Lemma 3 monogons and bigons can only bound regions that contain some special point; furthermore, a potential monogon about an elliptic point collapses — the geodesic passes through the elliptic point. To elucidate the application of these facts in the setting of extending finite geodesic configurations by rays, we first treat the following lemma.

Definition 4. A *cusped disk* on a hyperbolic surface or orbifold is a region that is bounded by a monogon and contains one cusp (and no other special points). An *internal* cusped disk is one for which the arms of the boundary monogon point into the cusped disk, see Figure 3.

Lemma 7. *Let α be a geodesic arc on any hyperbolic surface or orbifold. If α bounds an internal cusped disk, then α is not a highest arc of its geodesic.*

Proof. If either of the arms of α eventually ends in the cusp, then clearly our conclusion is met. We thus assume that this does not occur. See Figure 3 for the cases that we consider.

(A) If either geodesic arm of the monogon exits the cusped disk before meeting itself or the other arm, then a bigon bounding a region void of special points is formed, but these do not exist. If an arm meets itself first, then it either forms (B1) a monogon bounding a region void of special points, thus non-existent, or it creates (B2) a new cusped disk; but, the monogon bounding this cusped disk is higher than α. Finally, if the arms first meet each other, then they: (C1) meet so as to form a closed loop bounding a region void of special points, this monogon cannot exist; (C2) form a monogon about the cusp, but this newly formed loop and the original monogon about the cusped disk then give a bigon bounding a region void of special points, an impossibility; (C3) they form a bigon bounding a region void of special points, again impossible; or, finally, (C4) they form a bigon about the cusp, here any extension of either arm results in an impossible configuration. □

5. The 8-Fold Way of Low Height Arcs

Although there are many possible topological configurations of curves on $\Gamma^3\backslash\mathcal{H}$, Theorem 1, see §5.1, states that geodesic arcs of height between 3 and 6 can be distributed into eight of these; see Figure 4. The proof proceeds as a case-by-case analysis of configurations formed by oppositely oriented geodesic rays with common high point.

We say that a geodesic configuration is *about* some set of special points if this set is contained in a connected component of the complement of the configuration, compare with Definition 1.d.

Definition 5. Let α be a geodesic arc on $\Gamma^3\backslash\mathcal{H}$. We say that α is of one of the following types if, up to possibly changing its orientation, it has the corresponding properties:

- Type 1. The arc α is a pointed monogon about the cusp.
- Type 2. The arc α is a geodesic whose path has a single self-intersection, occuring at the node of an unpointed monogon about the cusp.
- Type 3. The arc α is a rho, emanating from an elliptic point, with highest point on its stem, and its monogon about the remaining two elliptic points; α has one point of self-intersection.
- Type 4. The arc α forms a monogon about the cusp, this monogon meets in its node a bigon about a single elliptic point; α has two points of self-intersection.
- Type 5. The arc α is the union of a rho whose monogon is about the cusp, with a rho whose monogon is about two elliptic points; α meets itself only at the monogons' nodes.

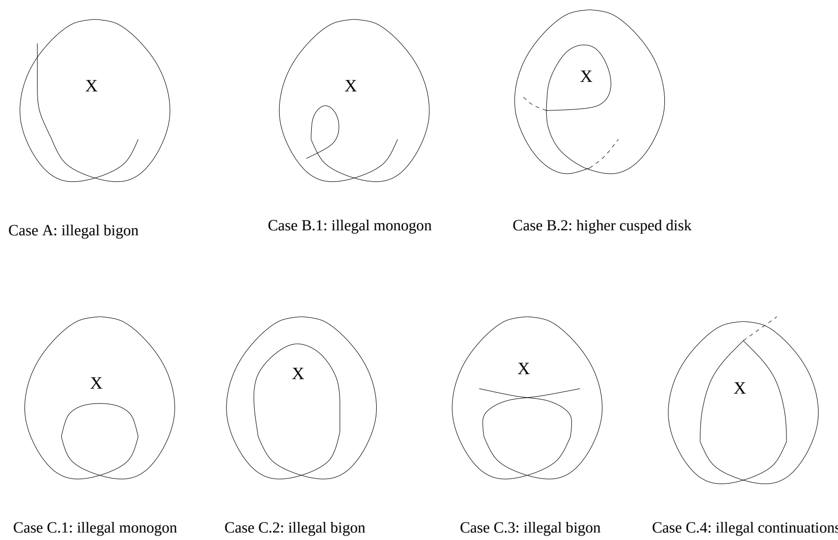

FIGURE 3. Internal cusped disks are not highest.

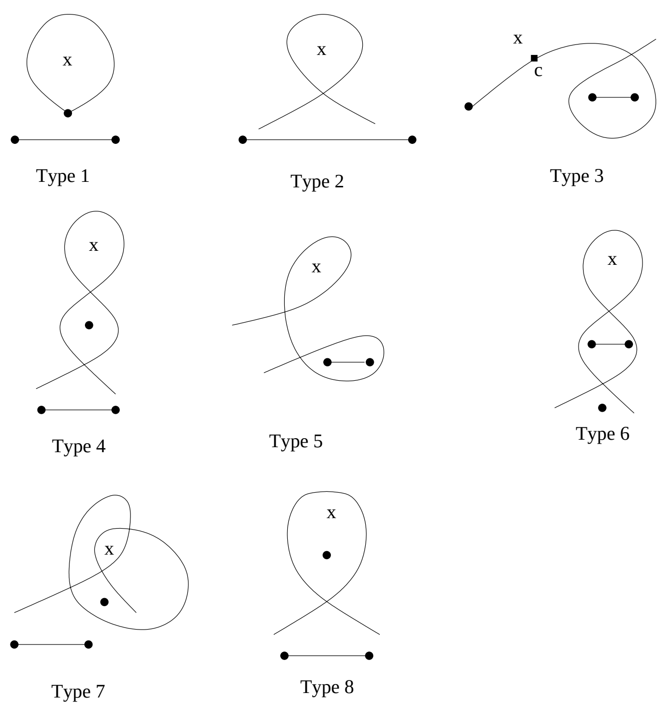

FIGURE 4. Topology of low nonsimple geodesics of $\Gamma^3\backslash\mathcal{H}$, with parameter simple closed geodesics. Cross denotes cusp; dots denote elliptic points; c is a highest point. (Variation of orientations and relative positions possible.)

Type 6. The arc α is the union of a monogon about the cusp, and a bigon about two elliptic points, these share a common node; α has two points of self-intersection.

Type 7. The arc α forms a monogon about the cusp, as well as a bigon about the cusp, and a bigon about one elliptic points; the arc α has three self-intersections.

Type 8. The arc α forms a monogon about the cusp and one elliptic point; α has one point of self-intersection.

Lemma 8. *If α as above is of any type other than Type 2, then there is a unique simple closed geodesic of $\Gamma^3\backslash\mathcal{H}$ that does not meet α.*

Proof. If α is of Type 1, 3, 5, 7, or 8, the definition of this type uniquely identifies a monogon bounding two elliptic points; if α is of Type 6, then the bigon contains two elliptic points; finally, if α is of Type 4, then there are two elliptic points in the region bounded by the union of the monogon and bigon. In all these cases, there is a curve connecting these elliptic points, which is unique up to homotopy. The result follows. □

Definition 6. For α as in Lemma 8, we call the associated simple closed geodesic the *parameter simple closed geodesic* of α, see Figure 4.

Remark 1. Candidate parameter simple closed geodesics for closed Type 2 geodesics can be imagined as in Figure 4, see also [SS3].

Definition 7. If γ is a height-achieving geodesic of $\Gamma^3\backslash\mathcal{H}$ and a highest subarc α of γ is of some Type n, then γ is of *Type n identified by α* and the high point of α is called a *defining high point.*

Remark 2. A geodesic of $\Gamma^3\backslash\mathcal{H}$ may achieve its height more than once, and thus perhaps be of more than one type. However, a result of Gbur [G], see Theorem 2 of Chapter 5 of [CF], implies that a geodesic that achieves its height more than six times must be a closed geodesic.

5.1. **Distribution into Type.**

Theorem 1. *Let γ be a height-achieving geodesic on $\Gamma^3\backslash\mathcal{H}$, of height greater than* 3 *and less than* 6*. Then γ is of (at least) one of the eight types given in Definition 7.*

Proof. Since γ is height-achieving, we can choose a highest point, c, of γ. Since γ is of height greater than 3, c is certainly not an elliptic point. We consider two geodesic rays emanating from c lying along the path of γ: ρ_+, whose positive direction is that of γ; and ρ_- of the opposite direction.

Since γ is non-simple, it suffices to consider the four cases of configurations of subarcs given by Lemma 6 on page 6. (In parentheses and italics we give brief phrases to aid in visualizing the geometry of the various cases.)

Case 1 (Two Rho). By Lemma 2, c is not the node of either of the monogons, thus by construction c lies on neither monogon. The two monogons divide $\Gamma^3\backslash\mathcal{H}$ into three regions: two each bounded by a single monogon and the third having the two monogons as its boundary. For each monogon, the cusp and c lie in the same connected component bounded by the monogon: otherwise there is a high point that is closer to the cusp than is c. Now, by Lemma 3 (a) and (c), for each monogon the connected component without the cusp must contain at least two elliptic points. Since there are only three elliptic points, this is an impossibility.

Case 2 (One Rho, Other Ray Stops at Elliptic Point). As for Case 1, the cusp and c must be in the same component bounded by the monogon of the rho. The terminus of the second ray is of course also in this component. Hence, by Lemma 3, the other component must contain the two remaining elliptic points. The geodesic is thus of Type 3.

Case 3 (Rays Meet). In this case, a monogon containing c is formed. If the point of intersection is an elliptic point, then γ is of Type 1. We now suppose otherwise. The region bounded by the monogon containing the cusp must contain at most one elliptic point, for otherwise the other region bounded by the monogon would collapse.

(a) If the component containing the cusp also contains an elliptic point, then γ is of Type 8.

(b) We thus now assume that there is no elliptic point within this component, thus it is a *cusped disk*.

(i) If the arms of the cusped disk continue internally to the cusped disk, Lemma 7 applies.

We may thus assume that the *arms continue externally to the cusped disk*.

Arms Do Not Return to Cusped Disk:

(ii) Assume neither arm continues so as to return to the monogon before meeting itself or the other arm. Then if the arms never meet one another, γ is of Type 2; otherwise Lemma 6 applied to these arms gives four subcases. The first subcase is (A) the arms extend simply from the node until forming monogons, with the two extensions meeting one another only at the node of the original monogon. However, there are only three elliptic points; as in Case 1, some component of a monogon must in fact collapse. If (B) one arm forms a monogon and the second ends in an elliptic point, then this monogon must contain both of the remaining elliptic points, and γ is thus of Type 5. The arms may meet (C), to form a bigon, or a monogon (if they meet so as to complete a closed geodesic). If the component not containing the cusp of this newly formed configuration were to contain all three elliptic points, then a double loop about the cusp would arise; by Lemma 5 this is a contradiction to γ being of height less than 6. Thus, this component contains either one elliptic point and γ is of Type 4, or two elliptic points, and γ is of Type 6.

If (D) one of the arms is an infinite simple ray and the other forms a monogon not meeting this ray, then the monogon bounds two regions. The cusp is in one region; by our hypothesis, the infinite simple ray lies in this region (indeed, so does all of the original cusped disk). Let us label the other region bounded by this monogon as R; by Lemmas 3 and 5, R must contain exactly two elliptic points. Since R contains simple closed loops about these two elliptic points, there is a simple closed geodesic (joining the two elliptic points) that lies completely within the region R. It is easily verified that this is the unique simple closed geodesic that does not meet the monogon.

By Lemma 4, the infinite simple ray either (1) spirals into a closed simple geodesic joining two elliptic points, or else (2) is the limit of rays, with each lying along the path of a simple closed geodesic. Now, if the infinite ray is of the first type and spirals into the simple closed geodesic that does not meet the monogon, then it comes arbitrarily close to this simple closed geodesic. Thus this infinite ray must enter R; in particular, it must meet the monogon. Similarly any infinite ray spinning about some other simple closed geodesic certainly meets the monogon. Finally, if a simple ray is of the second type, then by Corollary 2 it comes arbitrarily close to at least two elliptic points; it certainly meets the monogon. Therefore, in all cases these arms in fact do meet. Thus, this subcase is empty.

Arm Meets Cusped Disk: (iii) Finally, suppose an arm meets the cusped disk as its next intersection. The arc returning to the cusped disk is the extension of an arc beginning at the point of return: a monogon is formed. (A) If the point of return is not the node of the original cusped disk, then there is a remaining arc of the monogon about the cusped disk that forms a bigon with the returning geodesic arc (see the representation in Figure 4 of an arc of Type 7). We say that the monogon and bigon are *formed by our arm.* Lemma 3 implies that for each, both of the regions bounded by it must contain at least one special point, with those bounded by the monogon not containing solely one elliptic point. Thus, one region bounded by the monogon formed by our arm contains two elliptic points, the other contains the cusp and the remaining elliptic point, say e; the bigon must be about e.

Our arm cannot end at the cusp, thus it must extend so as to exit from the cusped disk; there is thus an arc of the arm crossing the cusped disk. This crossing arc divides the cusped disk into regions bounded by a bigon and a triangle. The bigon must contain a special point; thus it contains the cusp. Now, the crossing arc either (1) exits the cusped disk along the bigon or (2) along the monogon formed by our arm. In subcase (2), this crossing arc divides the monogon so as to create a new monogon and a bigon. One of the regions bounded by this new monogon contains the bigon formed by our arm and the triangle of the cusped disk — it thus contains exactly one special point, the elliptic point e, a contradiction. Thus, only subcase (1) pertains; we conclude that α is of Type 7.

Finally, we must consider (B) the possibility that the arm returning to the disk does so at the node. In this case, an obvious monogon is created by this arm. But, there is a continuous arc formed by the original monogon bounding the cusped disk and this new arc, this is also a monogon. We thus have two regions bounded by (new) monogons into which three elliptic points must be distributed. One of these components is forced to have at most one elliptic point in it; this is a contradiction to Lemma 3, and we conclude that this case is void.

Case 4 (One Rho, Other Ray Infinite Simple). The remaining case arises by a first ray continuing infinitely and simply from c while the second has an initial subarc meeting itself; these meet only at c. The second ray gives rise to a monogon, as in Cases 1 and 2 above, the component of the initial monogon not containing c must contain two elliptic points. We thus argue as for Subcase 3.b.ii.D above; this case is void.

We have now exhausted all possibilities of γ, and have thus proven our result. □

5.2. **Type 8 Arcs Not Highest.** Recall that a highest arc of a geodesic is an arc whose highest point is at least as high as the high point of any other arc of the given geodesic. We first show that geodesic arcs of a general "internal" Type 8 description cannot be highest arcs.

Lemma 9. *Let α be a geodesic arc on a hyperbolic orbifold. Suppose that α consists of a monogon about a cusp and an elliptic fixed point of order two, and that the geodesic arms that continue the monogon immediately enter this component. Then α is not a highest arc (with respect to this cusp) of a finite height geodesic.*

Proof. We discuss cases similar to that of the proof of Lemma 7; in particular, we can and do assume that no arm ends at the cusp.

(A) If an arm exits the disk bounded by the monogon (before meeting itself or the other arm), then the disk is partitioned into two components, bounded respectively by a monogon and a bigon. Neither of these components can be void of special points, nor can the component bounded by the monogon contain only the distinguished elliptic point. Thus, the only admissible configuration has the monogon forming a cusped disk and the bigon bounding a region containing the distinguished elliptic point. The second arm must continue into the newly formed *cusped disk*. This arm either first intersects itself, forming a new cusped disk, whose bounding monogon is higher than the original monogon; or, this second arm exits the cusped disk, resulting in a new bigon and a new triangle. Both components of the bigon must contain some special point(s), thus the cusp must lie in one of its components. But, this bigon is then higher than the original monogon.

(B) If an arm meets itself first, then it forms a monogon. This monogon must have some special point in it. If it contains the cusp, then this arm gives a higher arc; if the monogon contains both the cusp and the elliptic point, then an annulus void of special points is formed, this collapses. If the monogon contains only the internal elliptic point, this monogon collapses so that the arm ends at this special point. The only admissible continuation of the second arm, that does not create a higher arc than the original monogon, is to also end at this elliptic point — this is subcase C3 below.

If the arms first meet each other, then they: (C1) form a loop, thus a monogon, (C2) form a bigon, or (C3) both end at the elliptic point. In subcase C1, the monogon must contain a special point. If the monogon contains only the cusp, or the cusp and the elliptic point, then a higher arc has been found; if the monogon contains only the elliptic point, then the monogon collapses, this situation is thus in subcase C3.

In subcase C2, our original disk is partitioned into two components, bounded by a triangle and by the bigon respectively. To avoid the formation of a higher arc, the cusp must be in the interior of the triangle; to ensure a legal configuration, the distinguished elliptic point must be in the component bounded by the bigon. Now, both arms must extend to meet the bounding monogon of the original disk, or else a higher arc is found. But, these extended arms result in a partition of the original disk into at least three components (if the arms do not exit at the same point of the bounding monogon, then four components result). There are only two special points to be distributed into these components, certain unions of which are bounded by (collapsing) monogons or bigons. A contradiction is reached.

In subcase C3, either: (a) the curve joining the elliptic point to itself is homotopic to a simple closed curve joining this special point to itself, or (b) the two arms form a loop about the cusp (with the distinguished elliptic point on this loop). In the first subcase the geodesic is simple, and hence α does not extend to a geodesic with arms continuing into the monogon, contrary to hypothesis. In the second, the new loop is closer to the cusp than is α; therefore, α is not a highest arc. □

Proposition 1. *A geodesic arc of Type 8 is not the highest arc of its geodesic.*

Proof. Let α be a geodesic arc of Type 8. By Lemma 9, we need only consider the case of *external* continuation of the arms of a monogon bounding a region containing both the cusp and one elliptic point. (A) If an arm returns to the cusped disk before meeting itself, or any elliptic point (irrespective of meeting the other arm), then just as in Subcase 3.b.iii.A of the proof of Theorem 1, both a new bigon and a new monogon are formed. Once again by Lemma 3, the non-cusped region bounded by the new monogon requires both of the remaining special points; but the non-cusped region bounded by the bigon requires at least one of these. Thus, in all choices, there is a collapse. We conclude that neither arm can return to intersect the initial monogon unless it first meets itself.

(B) If an arm intersects itself before returning to the cusped disk, then a monogon is formed. One region bounded by this monogon includes the original cusped disk; to avoid collapse, the other must contain both of the remaining elliptic points. (But, then the original monogon and this new monogon both are monogons about these two elliptic points. If they have opposite orientations, they homotope to cancel each other. We thus may assume that they have the same orientation.) Now, since the second arm begins in a region devoid of special points, by Lemma 3, it cannot self-intersect before meeting at least one of these two monogons. As well, if the second arm forms a simple ray, then it must meet at least one of these monogons: any finite length simple ray has path ending at an elliptic point and by Corollary 2 any infinite simple ray comes arbitrarily close to some elliptic point. Thus, the second arm first certainly meets one or the other of these monogons before meeting itself. But, from Case A, it must be that this second arm meets the monogon about the two elliptic points (at a point besides the node). Fix this point of intersection, it divides the monogon into two arcs. Follow the second arm back from this point until it first meets the arc joining the nodes of the two monogons (this certainly exists, as this second arm begins at a point on this arc). This newly located arc and one of the two arcs of the monogon about the elliptic points completes a bigon bounding a region without special point, an impossibility. Therefore, this case is empty.

If the arms meet, then they may do so either by (C1) completing a single closed curve (forming a monogon), or else a bigon. However, the monogon collapses if it is about a single elliptic point; if it is about two elliptic points then the curve so completed homotopes to be a multiple of a simple closed geodesic (see [Sh]; one can also argue by lifting the curve to $\Gamma'\backslash\mathcal{H}$). If the two arms next meet so as to form a bigon, then this bigon must be about either one, or both of the two remaining elliptic fixed points. If (C2) this bigon is about both of these elliptic points, then the total arc so described — monogon and subsequent bigon — can be homotoped to be arbitrarily close to the path of the closed geodesic joining the two elliptic points. The geodesic either is simple, or else there is a higher arc. If (C3) the bigon is about exactly one of the two remaining elliptic points — see Figure 5, then at most one arm can have its path terminate at the third elliptic point. Indeed, if both do, then once again a closed curve homotoping to a multiple of a simple curve is formed. Furthermore, the bigon is not of the type reasonably called internal, for at least one arm would exit the bigon to cause an illegal bigon. Now, choose an arm whose path does not terminate at the third elliptic point. Since the arm neither meets itself, nor returns to the original monogon, before some other intersection arises, applying Corollary 2 we conclude that this arm returns to the bigon. Of course, this arm returns to the bigon by meeting the other arm; a new bigon is formed, it must be about the third elliptic point. Now consider the second arm, it continues externally from the bigon; indeed, this continuation is into a region devoid of special points bounded by the original monogon and the two bigons. Thus, this second arm must in fact meet at least one of these configurations. A first such intersection cannot be with the original monogon; but the first intersection being with either of the bigons would imply the existence of a new bigon, bounding a region devoid of elliptic points. Therefore, this subcase is empty.

(D) Finally, each ray may continue simply, without returning to the cusped disk. But, choose any loop about the cusp and the elliptic point that is interior to the cusped disk. There is a homotopy taking the finite region bounded by this loop arbitrarily close to the unique simple closed geodesic joining the remaining elliptic points. That is, the geodesic itself is simple. □

References

[BLS] A.F. Beardon, J. Lehner and M. Sheingorn, *Closed geodesics on a Riemann surface with application to the Markoff spectrum*, Trans. Amer. Math. Soc. 295 (1986), 635–647

[Bon] F. Bonahon, *Geodesic laminations on surfaces*, in "Laminations and foliations in dynamics, geometry and topology (Stony Brook, NY, 1998)", 1–37, Contemp. Math., 269, Amer. Math. Soc., Providence, RI, 2001.

[CEG] R. D. Canary, D. B. A. Epstein, P. Green, *Notes on notes of Thurston*, in " Analytical and geometric aspects of hyperbolic space (Coventry/Durham, 1984)", 3–92, London Math. Soc. Lecture Note Ser., 111, Cambridge Univ. Press, Cambridge, 1987.

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

[Cr] D.J. Crisp, *Closed geodesics on the punctured torus*, 1993 University of Adelaide Ph.D. dissertation

[CrEtAl] D.J. Crisp, S.Dziadosz, D.J. Garity, T. Insel, T.A. Schmidt, P. Wiles, *Closed Curves and Geodesics with Two Self-Intersections on the Punctured Torus*, Monatsch. fur Math, 125 (1998), 189–207.

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

[CM2] ———, *The Markoff spectrum and closed geodesics with one self-intersection on a torus*, Flinders University preprint, (1995)

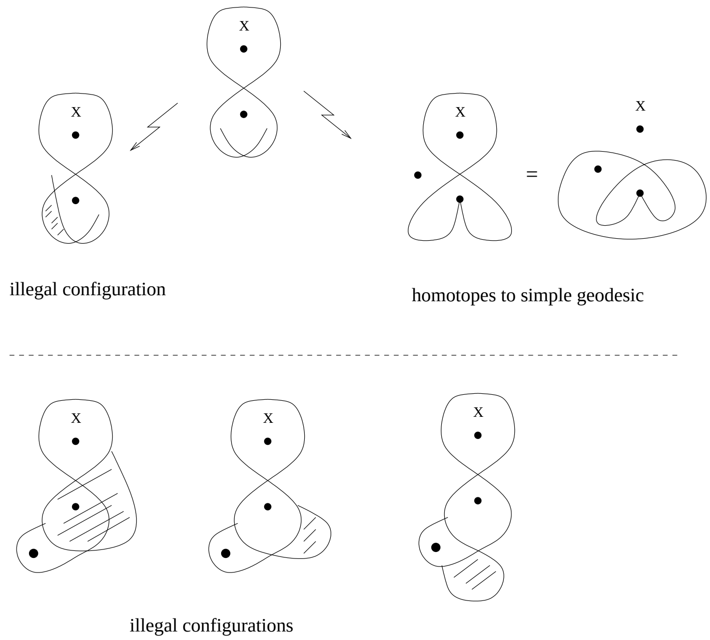

FIGURE 5. Type 8 not highest arc: Subcase C3.

[CF] T. W. Cusick and M. E. Flahive, "The Markoff and Lagrange spectra," Mathematical Surveys and Monographs 30, American Mathematical Society, Providence R.I., (1989)

[G] M. E. Gbur (now Flahive), *On the lower Markoff spectrum*, Monatsh. Math. 81 (1976), 95–107

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

[H2] ______, *Diophantine approximation on hyperbolic orbifolds*, Duke Math. J. 56 (1988), 531–547

[H3] ______, *Geometry Markoff Theory and a Theorem of Millington*, , in "Number theory with an emphasis on the Markoff spectrum," Lecture Notes in Pure and Applied Mathematics 147, A.D. Pollington and W. Moran, eds., Marcel Dekker, New York (1993), 107–112

[LS] 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

[Mc] ______, *Weierstrass points and simple geodesics*, Bull. London Math. Soc. 36 (2004), 181–187

[Sch] A. L. Schmidt, *Minimum of quadratic forms with respect to Fuchsian groups. I*, J. Reine Angew. Math. 286/287 (1976), 341–368

[SS1] T. A. Schmidt and M. Sheingorn, *Riemann surfaces have Hall rays at each cusp*, Ill. J. Math. 41 (1997), 378–397

[SS2] ______, *Parametrizing simple closed geodesy on* $\Gamma^3\backslash\mathcal{H}$, J. Aust. Math. Soc. 74 (2003), no. 1, 43–60

[SS3] ______, *Low Height Geodesics on* $\Gamma^3\backslash\mathcal{H}$ *: Explicit Height Formuli, Examples, and Local Isolation*, preprint

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

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

OREGON STATE UNIVERSITY, CORVALLIS, OR 97331
E-mail address: toms@math.orst.edu

CUNY - BARUCH COLLEGE, NEW YORK, NY 10010
E-mail address: marksh@alum.dartmouth.org