A remarkable identity for lengths of curves

Greg McShane

Thesis submitted for the degree of Doctor of Philosophy at the University of Warwick.

The Mathematics Institute,
University of Warwick,
Coventry.

May 1991

Acknowledgements

The motivation for the work lies in two series of seminars held at the University of Warwick. The first series concerned Thurston's unpublished work on minimal stretch maps between surfaces and the second Maskit's embedding of the Teichmuller space of the punctured torus. This thesis was written with the aid of Silvio Levy's LaTeX macros while visiting the Geometry Center (formerly the Geometry Supercomputer Project) in Minneapolis.

Principally, I wish to thank my supervisor David Epstein for his time and especially for his patience. His suggestions have been invaluable. The encouragements of Caroline Series and Troels Jorgensen are greatly appreciated. Troels's visits to Minnesota alone made my stay there worthwhile and Caroline's criticisms were crucial in formulating this work. I am also deeply indebted to Al Marden and all those at the Geometry Center for their assistance while preparing this thesis and for making my stay in Minneapolis a pleasant one.

Finally, I wish to thank SERC for their financial support.

Introduction

In this thesis we will prove the following new identity

$$\sum_{\gamma} \frac{1}{1+\exp|\gamma|} = \frac{1}{2},$$

where the sum is over all closed simple geodesics γ on a punctured torus with a complete hyperbolic structure, and $|\gamma|$ is the length of γ. Although it is well known that there are relations between the lengths of simple geodesics on a hyperbolic surface (for example the Fricke trace relations and the Selberg trace formula) this identity is of a wholly different character to anything in the literature. Our methods are purely geometric; that is, the techniques are based upon the work of Thurston and others on geodesic laminations rather than the analytic approach of Selberg.

The first chapter is intended as an exposition of some relevant theory concerning laminations on a punctured surface. Most important of these results is that a leaf of a compact lamination cannot penetrate too deeply into a cusp region. Explicit bounds for the maximum depth are given; in the case of a torus a simple geodesic is disjoint from any cusp region whose bounding curve has length less than 4, and this bound is sharp. Another significant result is that a simple geodesic which enters a *small* cusp region is perpendicular to the *horocyclic foliation* of the cusp region.

The second chapter is concerned with G_{cusp}, the set of ends of simple geodesics with at least one end up the cusp. A natural metric on G_{cusp} is introduced so that we can discuss approximation theory. We divide the geodesics of G_{cusp} into three classes according to the behaviour of their ends; each class also has a characterisation in terms of how well any member geodesic can be approximated. An example is given to demonstrate how this classification generalises some ideas in the classical theory of Diophantine approximation. The first class consists of geodesics with both ends up the cusp. Restricting to the punctured torus it is shown that for such a geodesic, γ, there is a portion of the cusp region surrounding each end which is disjoint from all other geodesics in G_{cusp}. We call such a portion a *gap*. The geometry of the gaps attached to γ is described and their area computed by elementary trigonometry. The area is a function of the length of the unique closed simple geodesic disjoint from γ. Next we consider the a *generic* geodesic in G_{cusp}, that is, a geodesic with a single end up the cusp and another end spiralling to a minimal compact lamination which is not a closed geodesic. We show that such a geodesic is the limit from both the *right* and *left* of other geodesics in G_{cusp}. Finally we give a technique for approximating a geodesic with a single end up the cusp and the other end spiralling to a closed geodesic. Essentially we repeatedly *Dehn twist* a suitable geodesic in G_{cusp} round this closed geodesic. The results of this chapter are then combined with a theorem of J. Birman and C. Series to yield the identity.

Contents

List of Figures

Chapter 1

Cusp Regions and Simple Geodesics

In this chapter we review some facts about geodesic laminations on a surface with a puncture. Much of this appears in some form in [Thu] but a better exposition of the material is given by [CEG]. We begin by giving some definitions.

1.1. Preliminaries

Definition 1.1.1. A *surface* is a two manifold together with a complete hyperbolic structure of finite area. A *punctured surface* is a surface with at least one cusp.

A *curve* on the surface is the image in the surface of a continuous and piecewise smooth mapping of the open unit interval. A curve is *simple* if it does not self intersect. A curve is *complete* if it is complete as a metric space with respect to the path metric induced on it from the surface.

A *geodesic* is a curve which is locally length minimising. A *lamination* is a collection of disjoint simple geodesics which is closed as a pointset. A non-empty lamination is *minimal* if no proper subset is a lamination.

Lemma 1.1.2 (classes of simple geodesic). *On a surface any complete simple geodesic falls into exactly one of the following three classes*

(a) It is a leaf of a compact lamination.

(b) It has a single end spiralling into a compact lamination and the other end up a cusp.

(c) It has both ends up a cusp.

This lemma appears with slightly different emphasis in [CEG].

Remarks: We note that a closed simple geodesic is a compact lamination and so falls into the first of the three classes. Further on a surface without punctures (and so without cusps) all simple geodesics fall into the first class.

1.2. Punctured Tori

A punctured torus is a surface homeomorphic to a torus with one point deleted. It is a well known fact (see for instance [Abi]) that a punctured torus has a complete hyperbolic structure and every such structure has area 2π.

In this section we use the basic technique of cutting a surface along certain curves to get a *cut surface.* Choose a disjoint set of simple curves for which the underlying pointset is closed as a subset of the surface. The complement of this set of curves in the surface is an open surface. The *cut surface* is the metric completion of this open surface with respect to its path metric.

We classify the surfaces which can be obtained from a punctured torus by cutting along compact minimal laminations. Such a lamination may consist of a single closed geodesic or an uncountable collection of leaves. This dichotomy manifests itself when we form the cut surface. When we cut a punctured torus along a closed geodesic the surface we get is a pair of pants. There are compact minimal laminations on the punctured torus which are not closed geodesics, although they are the limits of closed simple geodesics (in the Hausdorff topology on closed subsets). The surface we obtain when we cut along such a lamination is not a pair of pants but a punctured ideal bigon.

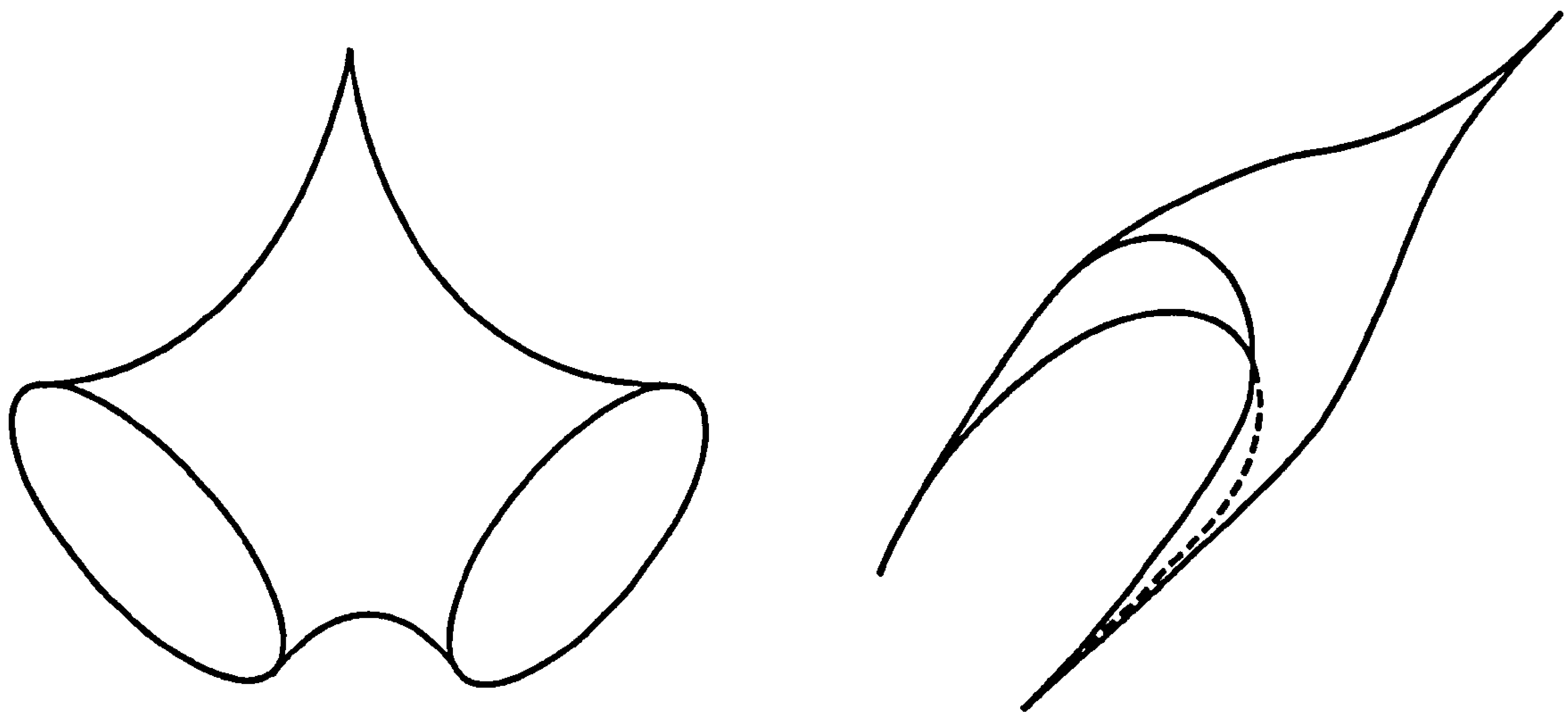

Figure 1.1. The two types of cut surface obtained from a punctured torus. The surface on the right is a pair of pants. The surface on the left is a punctured ideal bigon. The tips of the spikes are in fact ideal vertices.

Lemma 1.2.1 (cut lamination not a closed geodesic). *If a punctured torus is cut along a minimal compact lamination which is not a closed geodesic then the cut surface is a punctured ideal bi-gon. That is, it is isometric to an ideal triangle doubled along two of its edges.*

Proof of 1.2.1: Let β be a minimal lamination which is not a closed geodesic. We can choose a sequence, γ_i, of closed geodesics which converge to β in the

Hausdorff topology, so that the lengths are monotone increasing. We cut along γ_i to obtain a sequence of pairs of pants.

We choose a point on the torus disjoint from all simple closed geodesics; such a point exists since the union of all simple closed geodesics is a set of measure zero on the surface. The image of this point on each of the pants is a base point. The sequence of pants with this base point converges to the double of an ideal triangle in the geometric topology (the lengths of the boundary components increases monotically and the length of the simple geodesic arcs joining these boundaries decreases monotically.) 1.2.1

Lemma 1.2.2 (minimal laminations intersect). *Two distinct minimal compact laminations on a punctured torus intersect.*

Proof of 1.2.2: Let γ be a minimal compact lamination. We cut along γ and analyse the cut surface we obtain. If the torus has a lamination disjoint from γ then there must be a corresponding lamination on this cut surface.

Suppose γ is a closed geodesic. Cutting along γ we obtain a pair of pants. The only closed simple geodesics on a pair of pants are contained in the boundary; that is, there are no closed geodesics other than those corresponding to γ. So there are no simple closed geodesics on the torus which are disjoint from γ. In addition the simple closed geodesics are dense in minimal laminations; so the cut surface has no minimal laminations other than those contained in the boundary. Thus there are no minimal laminations disjoint from γ.

Suppose γ is not a closed geodesic. When we cut along it now the cut surface is a punctured ideal bi-gon. This surface has no closed simple geodesics and so can have no minimal laminations. 1.2.2

Decomposing into Quadrilaterals

In order to perform certain calculations we find it convenient to decompose the punctured torus into four congruent quadrilaterals. Each quadrilateral has angles $0, \pi/2, \pi/2, \pi/2$ at its vertices. This decomposition depends on the choice of a closed simple geodesic.

Choose a closed simple geodesic and cut the surface along it. The resulting surface is isometric to a degenerate pair of pants; that is, the cut surface has two geodesic boundary components and a single cusp. For each of these geodesic boundary components there is a unique simple geodesic running from the cusp and meeting the boundary perpendicularly. There is also a unique simple geodesic which meets both boundary components perpendicularly. These three perpendiculars do not intersect. Cutting along the perpendicular geodesics we obtain a congruent pair of pentagons each with a single ideal vertex.

We take a pentagon and cut along the geodesic which runs from the ideal vertex to meet the opposite edge perpendicularly. This gives a pair of quadrilaterals each with a single ideal vertex.

Since we always cut along curves which met the boundaries of these shapes perpendicularly, all the angles in the quadrilateral, with the exception of the ideal vertex, must be $\pi/2$.

To make our calculations it is useful to identify a quadrilateral of the type obtained above with a congruent quadrilateral in a standard position in the upper half plane model. Such a quadrilateral is illustrated in figure 1.2.

Lemma 1.2.3 (standard position for quadrilaterals). *Any quadrilateral with vertex angles $0, \pi/2, \pi/2, \pi/2$ is congruent to a quadrilateral in $\boldsymbol{H}$ such that*

(a) the ideal vertex is the point at infinity in $\boldsymbol{H}$;
(b) the two sides of the quadrilateral which meet at the ideal vertex are contained in the line $\{z : Re\, z = 0\}$ and the line $\{z : Re\, z = 1\}$.

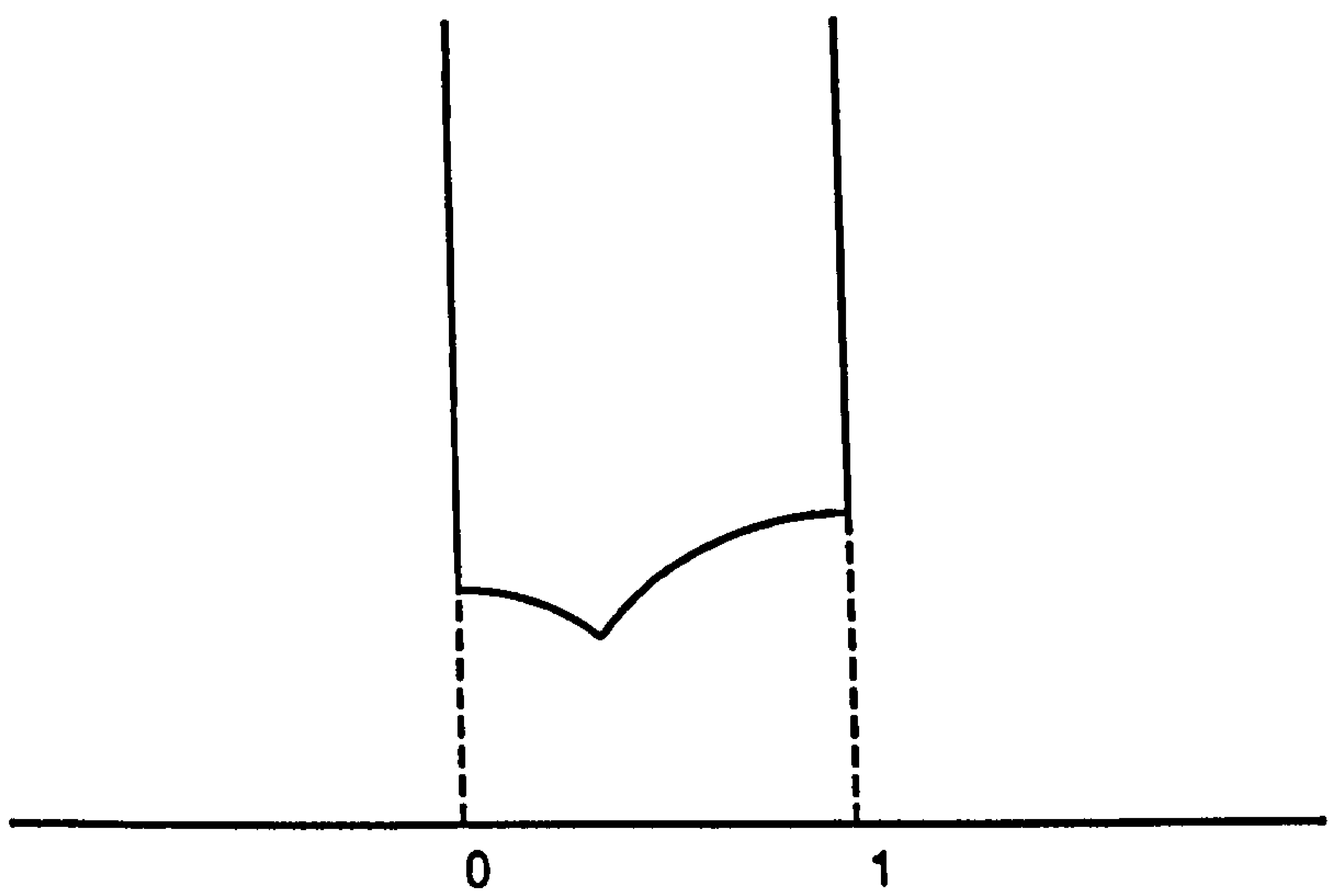

Figure 1.2. A quadrilateral in standard position. This is a quadrilateral in standard position in the upper half plane model. There are only three vertices shown; the fourth vertex is ideal and is the point at infinity.

1.3. Behaviour of Geodesics with Respect to Cusp Regions

We define a cusp region and investigate the relationship between simple geodesics and small cusp regions. In particular we show leaves of compact laminations cannot enter small cusp regions and that a simple geodesic must behave in a very nice way in a sufficiently small cusp region.

Definition 1.3.1 (cusp region). A *cusp region* is a portion of the surface isometric to $\{z : \operatorname{Im} z \geq 1\}/[z \mapsto z + p]$. The length of the horocyclic curve bounding the cusp region is p.

Convention:
In what follows we will specify a cusp region by the length of the horocyclic curve bounding the region. By a small cusp region we mean one for which the bounding curve is short.

The following lemma shows that minimal compact laminations cannot go too far up the cusp; that is, they cannot enter very small cusp regions in the sense of the convention above. This particular lemma applies only to punctured tori. The lower bound on the length of the bounding curve is too big for a higher genus surface.

Lemma 1.3.2 (compact laminations on punctured torus). *Let $\epsilon > 0$.*

Any punctured torus has a cusp region with bounding curve of length $4-\epsilon$. This is the best possible lower bound for the size of the largest cusp region over the moduli space of punctured tori.

Any minimal compact lamination on a punctured torus does not intersect the cusp region whose bounding curve has length $4-\epsilon$.

Proof of 1.3.2:

Any minimal compact lamination is the limit, in the Hausdorff topology, of simple closed geodesics. Therefore it is sufficient to show that no simple closed geodesic enters the cusp region with bounding curve of length 4 (the $4-\epsilon$ is the result of blunting of the inequality when we take limits).

Take any simple closed geodesic, γ, and decompose the surface into four quadrilaterals as described in the previous section. Each of these is congruent to the same quadrilateral in standard position in the upper half plane.

This quadrilateral has two sides of finite length neither of which intersect the line $\{z : \operatorname{Im} z = 1\}$. To see this note that these sides lie on a pair of orthogonal circles with centres a distance 1 apart, so each circle has a radius less than 1. Since the centres of these circles lie on the real line no point on either circle can have imaginary part bigger than or equal to 1.

The portion of $\{z : \operatorname{Re} z = 1\}$ interior to the quadrilateral has length 1. Since each quadrilateral in the decomposition of the torus is congruent to this one there is a corresponding horocyclic curve on each of them.

We re-glue the quadrilaterals to obtain a surface isometric to the original torus. The union of these four curves is a closed horocyclic curve bounding a cusp region on this torus; this curve has length 4. Since each of the four horocyclic curves is disjoint from the finite sides of the quadrilaterals this bounding curve is disjoint from the closed geodesic γ which we chose. We note that because the cut surface corresponding to certain laminations is a punctured ideal bigon this is the best possible bound.

1.3.2

Simple Geodesics which Meet Small Cusp Regions

We now describe the behaviour of simple geodesics after they have entered some small cusp region. To do this we introduce the horocyclic foliation of the cusp region which, for suitably small cusp regions, a simple geodesic must always intersect orthogonally.

Definition 1.3.3 (Horocyclic Foliation). Let C be a cusp region. We can foliate C by geodesics with an end up the cusp and which meet the curve that

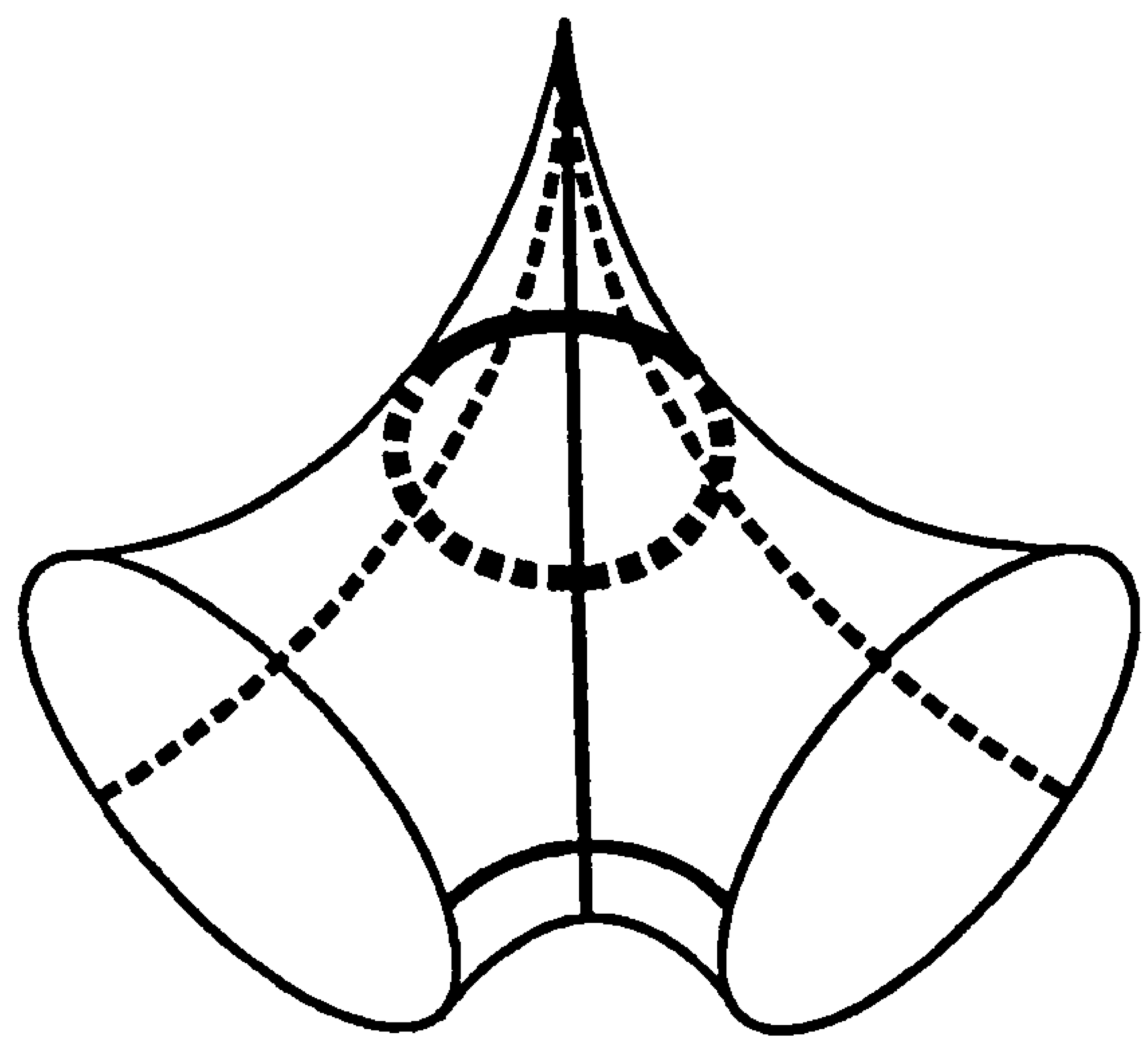

Figure 1.3. Four quadrilaterals glued to make pants. The four quadrilaterals have been glued together to make a pair of pants. By identifying the boundary geodesics suitably we can obtain a surface isometric to the original torus. The heavy circle marked on the pants is the horocycle of length 4.

bounds C perpendicularly. The *horocyclic foliation* is the foliation orthogonal to this.

Remarks:

The leaves of this foliation are concentric horocyclic curves homotopic to the cusp point.

If a cusp region is included in a larger cusp region then each leaf of its horocyclic foliation is a leaf of the horocyclic foliation on the larger one.

Lemma 1.3.4 (behaviour high up cusp). *Let M be a surface with a single cusp; such a surface has a cusp region of length* 2. *The portion of a complete simple geodesic which intersects a cusp region whose bounding curve has length less than* 2 *always meets the horocyclic foliation perpendicularly.*

Proof of 1.3.4: We see that there is a cusp region of length 2 as follows. There is a pair of disjoint simple geodesics, α and β, such that the component of $M\backslash(\alpha \cup \beta)$ containing the cusp is isometric to a pair of pants. This pair of pants can be decomposed into 2 similar pentagons as we did with the punctured torus as follows. We cut the pants along the two simple geodesics which have an end up the cusp and meet the boundary geodesics perpendicularly and along the simple geodesic perpendicular to both boundary geodesics. The resulting surface is a pair of pentagons each with four right angles and a single ideal vertex. We can identify these with a pentagon in the upper half plane which has its infinite sides contained in the lines $\{z : \operatorname{Re} z = 0\}$ and $\{z : \operatorname{Re} z = 1\}$. The finite sides of this pentagon are arcs of three semicircles each of which has its centre on the real axis; these semicircles are orthogonal because the pentagon had 4 right angles. The semicircle which intersects $\{z : \operatorname{Re} z = 0\}$ meets it at a right angle and so must have its centre at 0. Similarly the semicircle which meets $\{z : \operatorname{Re} z = 1\}$ has its centre at 1. These two semicircles cannot intersect because they meet the third semicircle orthogonally, hence they must have radius less than 1. Also the third semicircle must have radius less than 1 because it meets these orthogonally. So, since the three semicircle have radius less than 1, they do not intersect the line $\{z : \operatorname{Im} z = 1\}$. The portion of this line interior to the pentagon has length 1. When we glue the pentagons back together to get a pair of pants isometric to the original the images of these line segments is a horocycle of length 2.

We conjugate the covering group of the surface M so that the cusp region whose bounding curve has length 2 is covered by $\{z : \operatorname{Im} z \geq 1\}$, $[z \mapsto z+2]$. The horocyclic foliation lifts to a foliation of $\{z : \operatorname{Im} z \geq 1\}$ each leaf of which is a horocycle based at the point at infinity in the upper half plane.

We can pick a lift of the geodesic to $\mathbf{H}$ so that some portion lies above the line $\{z : \operatorname{Im} z \geq 1\}$. This lift is either a Euclidean semicircle of diameter greater than 2, or a vertical line. The lift cannot be a semicircle. If it were a circle its image under $z \mapsto z+2$ would still be a lift of the geodesic. This new lift would intersect the original because both are circles of diameter greater than 2. This contradicts the hypothesis that the geodesic was simple on the surface.

Thus the lift has to be a vertical line and the geodesic must meet the horocyclic foliation on the cusp region orthogonally.

1.3.4

An immediate consequence of this lemma is the following analogue of lemma 1.3.2.

Corollary 1.3.5 (compact laminations on other surfaces). *Let $\epsilon > 0$. Any geodesic which is the leaf of a compact lamination on a punctured surface does not intersect the cusp region whose bounding curve has length $2-\epsilon$.*

Proof of 1.3.5: By the lemma a complete simple geodesic which intersects the cusp region whose bounding curve has length $2-\epsilon$ must have an end up the cusp. Thus it cannot be contained in any compact subset of the surface.

1.3.5

Corollary 1.3.6 (simple geodesics disjoint in cusp region). *Let γ be a complete simple geodesic on a punctured surface. The geodesic γ intersects no other complete simple geodesic in any cusp region whose bounding curve has length less than 2.*

Chapter 2

Ends of Simple Geodesics

2.1. A Metric for the Set of Ends of Simple Geodesics

Let M be a punctured surface. Let C be a cusp region whose horocyclic bounding curve, ∂C, has length less than 2. We now introduce a metric on the set of ends of simple geodesics on the surface M. There are two equivalent definitions of this metric. Both are natural in the sense that they come from the geometry of the surface and in particular from considerations of the cusp region C. Later this metric will be used to investigate the manner in which various classes of ends can be approximated by sequences of ends.

Definition 2.1.1 **($G_{\mathbf{cusp}}$).** The intersection of C with a complete simple geodesic is empty, or has one component, or has two components. We denote by G_{cusp} the set of all such components. This set can be identified with a certain closed subset of C.

Remark: A complete simple geodesic with two ends up the cusp will give rise to two points of G_{cusp}.

The set S of all points x, x in some γ, γ in G_{cusp}, is a closed subset of the surface M. This is because the set of geodesic laminations is closed in the space of all closed sets in the Hausdorff topology.

Defining the Metric

Let β and γ be geodesics in G_{cusp}.

We define the distance between β and γ to be the minimum distance measured by arclength along ∂C between the sets $\gamma \cap \partial C$ and $\beta \cap \partial C$, divided by the total length of ∂C.

This definition is independent of the choice of cusp region.

Definition 2.1.2 (strip). A *strip* is a component of the complement in the cusp region C of some (non-empty) closed collection of geodesics in G_{cusp}.

Any strip is congruent to the region $\{z : \text{Im}z \geq 1, 0 < \text{Re}\, z < p\}$ for some $p > 0$. This is a figure bounded on two sides by geodesics which meet in an ideal vertex and on the other side by a horocyclic curve based at this ideal vertex.

Let β and γ be two geodesics in G_{cusp}. We could define a distance between these geodesics by taking the minimum of the areas of the two strips which have one geodesic side contained in β and the other geodesic side contained in γ and then dividing by the area of the cusp region. The following lemma shows that this metric is in fact the same as the one defined earlier.

Lemma 2.1.3 (area of strip). *The area of a strip is just the length of its horocyclic side.*

Proof of 2.1.3: We calculate the area of the strip congruent to $\{z : \operatorname{Im} z \geq 1, 0 < \operatorname{Re} z < p\}$, $p > 0$. The length of the horocyclic side is p. The hyperbolic area element is $dx\,dy/y^2$, so the area of the strip is

$$\int_1^\infty \int_0^p \frac{dx\,dy}{y^2},$$

which is just p. [2.1.3]

2.2. Structure of the Ends on a Punctured Torus

In this section, with the exception of theorem 2.2.4, we fix our surface M to be a punctured torus. We show that there are non-empty strips in the cusp region which are disjoint from all geodesics in G_{cusp}. Each such strip is associated with a unique simple closed geodesic on the punctured torus. In section 2.5 we will show further that any such strip in the cusp region arises, in a way we will explain, from a closed geodesic. We will show that each end of a geodesic with both ends up the cusp is an isolated point of G_{cusp} and that all isolated points of G_{cusp} arise in this way.

Lemma 2.2.1 (disjoint from unique minimal). *On a punctured torus a simple geodesic with a single end up the cusp is disjoint from exactly one minimal lamination. The geodesic spirals into this lamination.*

Proof of 2.2.1: Let γ be a simple geodesic with a single end up the cusp. If γ intersects the minimal lamination in its closure then it must self intersect. So it is disjoint from at least one lamination. However, by lemma 1.2.2 any other compact lamination intersects this one. So γ must intersect every other compact lamination on the punctured torus. [2.2.1]

Lemma 2.2.2 (number of disjoint geodesics). *On a punctured torus:*

(a) A closed simple geodesic is disjoint from exactly one simple geodesic with both ends up the cusp. It is also disjoint from exactly four simple geodesics which each have a single end up the cusp.

(b) A compact minimal lamination which is not a closed geodesic meets every geodesic with both ends up the cusp. It is disjoint from exactly two simple geodesics which each have a single end up the cusp.

Proof of 2.2.2: We cut along the minimal lamination and examine the various closed geodesics on the cut surface.

Suppose the lamination is a simple closed geodesic. The cut surface is a pair of pants. There is a unique proper homotopy class of simple curves on the surface with both ends up the cusp and so there must be a unique simple geodesic with both ends up the cusp. Every closed simple curve on this surface is homotopic either to the cusp or to one of the boundary components. So there are no compact minimal laminations on the surface other than those contained in the boundary. Fix a boundary component. There are exactly two simple geodesics which each have an end up the cusp and another end spiralling to this boundary component. So there are four geodesics with a single end up the cusp disjoint from the geodesic we cut along.

Suppose the lamination is compact and minimal but not a closed geodesic. The cut surface is a punctured ideal bigon. There is no proper homotopy class of simple closed curve with both ends up the cusp on this surface. We think of the bigon as being the double of an ideal triangle along two of its edges. In this way we see that there are two simple geodesics each with a single end up the cusp, namely the geodesic edges of the triangle along which we doubled. 2.2.2

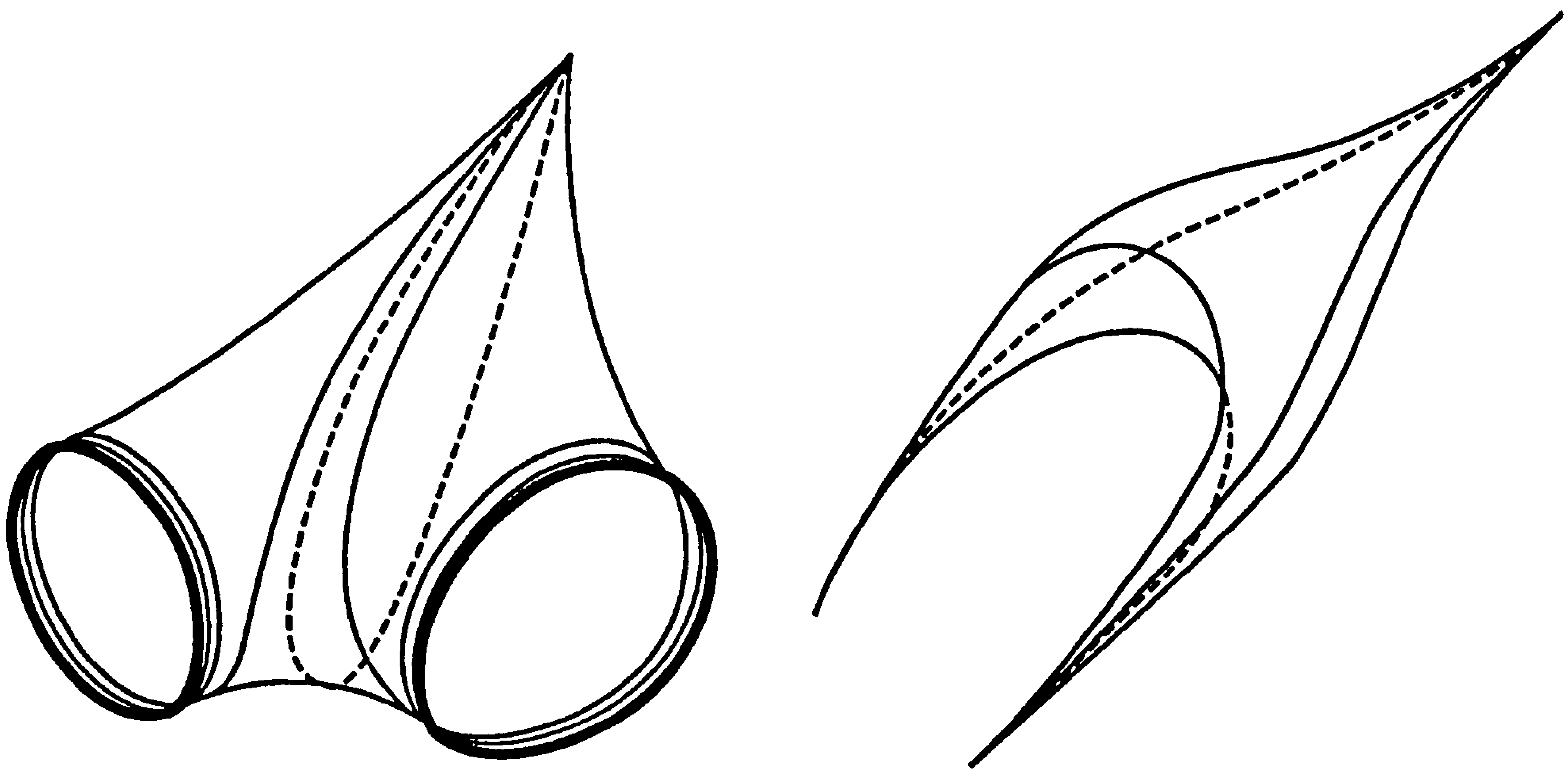

Figure 2.1. Complete geodesics on the cut surfaces. Geodesics on cut surfaces give rise to geodesics on the torus disjoint from the lamination along which the cut was made. Marked on the pair of pants there is a geodesic with both ends up the cusp and a pair of geodesics each with an end up the cusp and another end spiralling to the boundary. For clarity the two other geodesics with this property are not marked. Marked on the punctured bigon are the two geodesics each with an end up the cusp.

Proposition 2.2.3 (calculation). *Let M be a punctured torus. Let β be a simple geodesic with both ends up the cusp and α the closed simple geodesic disjoint from β. Choose a component of the complement of $\beta \cup \alpha$ in M and choose an end of β. Then there is a simple geodesic with a single end up the cusp, the other end spiralling to α, and lying entirely in the chosen component; such that the distance from β to this geodesic is exactly*

$$\frac{1}{2(1+\exp|\alpha|)}.$$

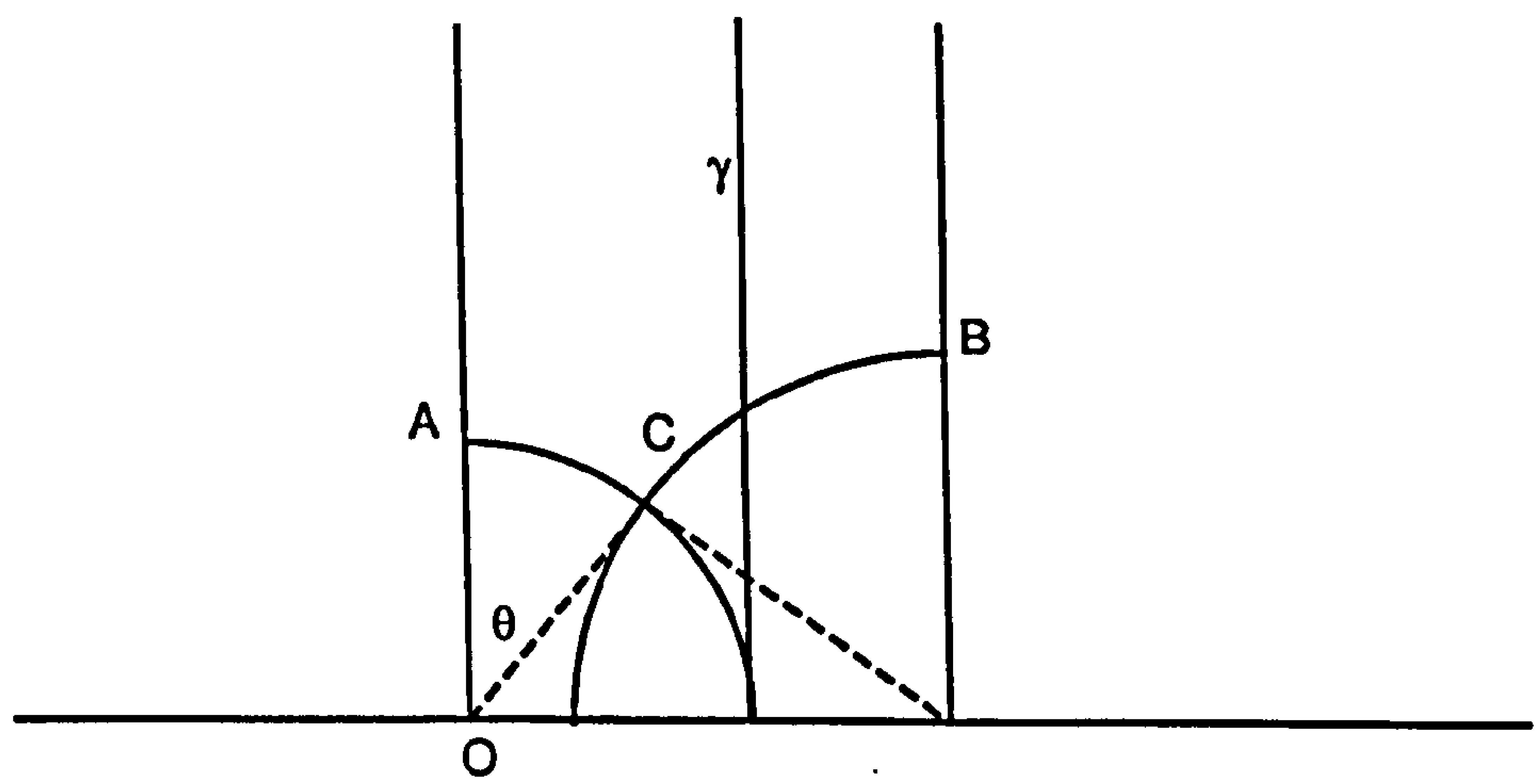

Figure 2.2. Quadrilateral in standard position. The figure represents a quadrilateral in standard position. The vertices A, B, C are marked as is the angle θ. The vertical line in the middle contains the end of the geodesic γ which goes up the cusp.

Proof of 2.2.3: We decompose the surface into four quadrilaterals as in section 1.2, with α as the chosen closed geodesic. The component of $M\backslash\alpha \cup \beta$ is the union of two of these quadrilaterals. Each quadrilateral has a single infinite side contained in β, and a single finite side contained in α. The length of this side is $|\alpha|/2$. We choose the quadrilateral which is contained in the appropriate component of $M\backslash\alpha \cup \beta$ and which has the appropriate end of β as a side. By lemma 2.2.2 there are exactly four geodesics each with an end up the cusp and another spiralling to the geodesic α on the torus. The end of such a geodesic which goes up the cusp is contained in a single quadrilateral. Let γ be the geodesic which spirals to α and whose other end is contained in the chosen quadrilateral.

We identify this quadrilateral with one in standard position in the upper half plane (lemma 1.2.3). This can be done in such a way that the side contained in β is gets mapped into $\{z : \operatorname{Re} z = 1\}$. With these choices a parabolic generator preserving ∞ is $z \mapsto z+4$.

Let A be the finite vertex of the quadrilateral lying on the side $\{z : \operatorname{Re} z = 0\}$ and B be the finite vertex of the quadrilateral lying on the side $\{z : \operatorname{Re} z =$

$1\}$. Let C be the other finite vertex of the quadrilateral. Let a be the Euclidean height of the point A above the real line, and b be the height of the point B.

The side AC is a portion of the geodesic α, and it is contained in a geodesic which has an endpoint at a. So the end of γ contained in the quadrilateral gets mapped into $\{z : \mathrm{Re}\, z = a\}$. The intersection of the horocycle of length 4 and the quadrilateral in the surface gets mapped into $\{z : \mathrm{Im}\, z = 1\}$. So distance, measured along the horocycle of length 4, between the end of γ and the end of β which is a side of the quadrilateral is $1 - a$.

Let θ be the angle between the Euclidean lines OA and OC. One sees that $\tan\theta = a/b$, $\sec\theta = 1/b$ and $a^2 + b^2 = 1$.

We compute the hyperbolic length of AC directly;

$$\begin{aligned}\text{length AC} &= \int_0^\theta \sec\theta d\theta \\ &= \log(\tan\theta + \sec\theta) \\ &= \log\frac{1+a}{b} \\ &= \frac{1}{2}\log\frac{1+a}{1-a}.\end{aligned}$$

On rearrangement this yields

$$1 - a = \frac{2}{1 + \exp(|\alpha|)}.$$

The value of $1 - a$ is 4 times the distance, in the metric on G_{cusp}, between the geodesic γ and the end of β which is a side of the quadrilateral.

2.2.3

Theorem 2.2.4 (spiralling-in closer). *Let M be a once punctured surface with finite volume and no boundary and let γ, β be geodesics in G_{cusp}. Let α be a closed simple geodesic which intersects γ and is such that β neither intersects α nor spirals into α . There is a geodesic in G_{cusp}, with a single end up the cusp nd the other end spiralling to α, which is nearer to β than γ is with respect to the metric on G_{cusp}.*

Proof of 2.2.4: We lift to H. Conjugate the covering group so that the horocycle of length 1 lifts to $\{z : \mathrm{Im}\, z = 1\}$. The geodesic α lifts to a lamination $\hat{\alpha}$.

Choose a lift $\hat{\gamma}$ of γ which is a vertical line. Travelling from the cusp at infinity, let F be the first leaf of the lamination $\hat{\alpha}$ which $\hat{\gamma}$ hits, and let α_1 and α_2 be the pair of vertical lines which end at the endpoints of F. The image of α_1 on the surface is a simple geodesic, since we can find a simple curve on the surface which with a lift to H which has the same endpoints as the lift of α_1. Similarly the image of α_2 on the surface is simple. Both of these geodesics

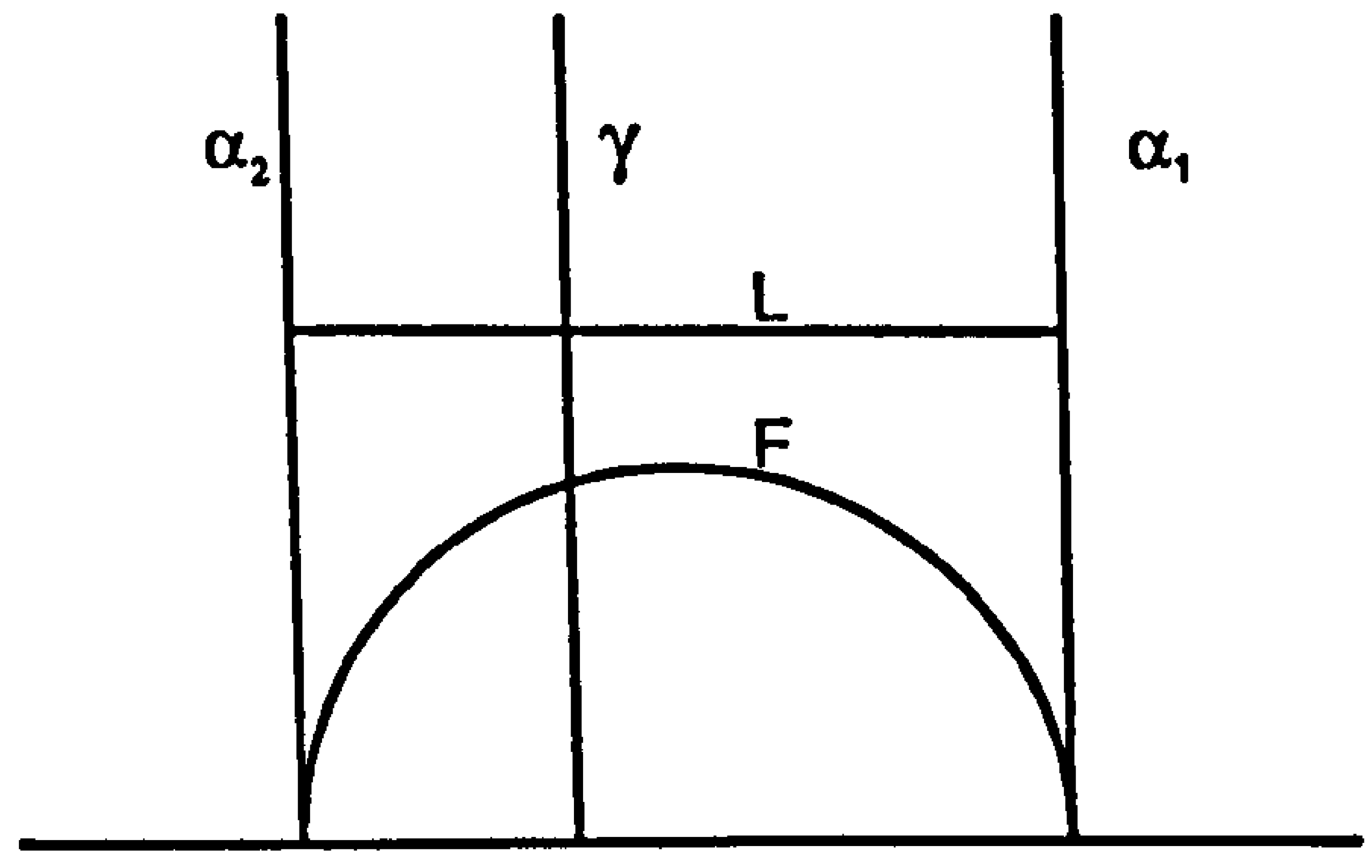

Figure 2.3. Diagram of lifts to H

spiral into α because they each share an endpoint with F. We show that one of α_1, α_2 is nearer to β than γ is.

The horocyclic segment L contained in $\{z : \operatorname{Im} z = 1\}$ which abuts on α_1 and α_2 maps injectively to a horocyclic segment for the following reasons. The line $\{z : \operatorname{Im} z = 1/2\}$ is a lift of the horocycle of length 2 on the surface. Since α is simple and closed it cannot intersect the horocycle of length 2, by corollary 1.3.5. So the lift F of α cannot intersect $\{z : \operatorname{Im} z = 1/2\}$. Therefore the Euclidean diameter of the semicircle F is less than 1. So L projects to something of length less than 1 on the horocycle of length 1 on the surface.

We will be done if no lift of β can intersect L. To see this note that α_1 is not a lift of β because the image of β on the surface cannot spiral into α. For the same reasons α_2 is not a lift of β. If there were a lift of β which met L then it would intersect L orthogonally by lemma 1.3.4. Hence it would meet F, which is impossible.

2.2.4

Theorem 2.2.5 (existence of gaps). *Let M be a punctured torus and let β be a simple geodesic with both ends up the cusp. There are a pair of strips such that β is the only simple geodesic which intersects either of these strips. The width of each such strip is*

$$\frac{1}{1+\exp|\alpha|}$$

where α is the unique closed simple geodesic disjoint from β.

Proof of 2.2.5: Let γ be a simple geodesic with an at least one end up the cusp and which does not spiral into α.

Choose an end of β. By theorem 2.2.4 we can find a pair of geodesics which each spiral into α and so that γ only intersects one of the strips in the complement and the end of β intersects the other. So one of the geodesics which spiral into α is closer to this end of β than γ is. It follows that the four geodesics with an end up the cusp and which each spiral into α bound a pair of strips so that β is the only simple geodesic to intersect either strip. By proposition 2.2.3 the closest any such geodesic gets, in the metric of G_{cusp}, to an end of β is

$$\frac{1}{2(1+\exp|\alpha|)}.$$

So the width of the strip containing the end of β is just twice this.

2.2.5

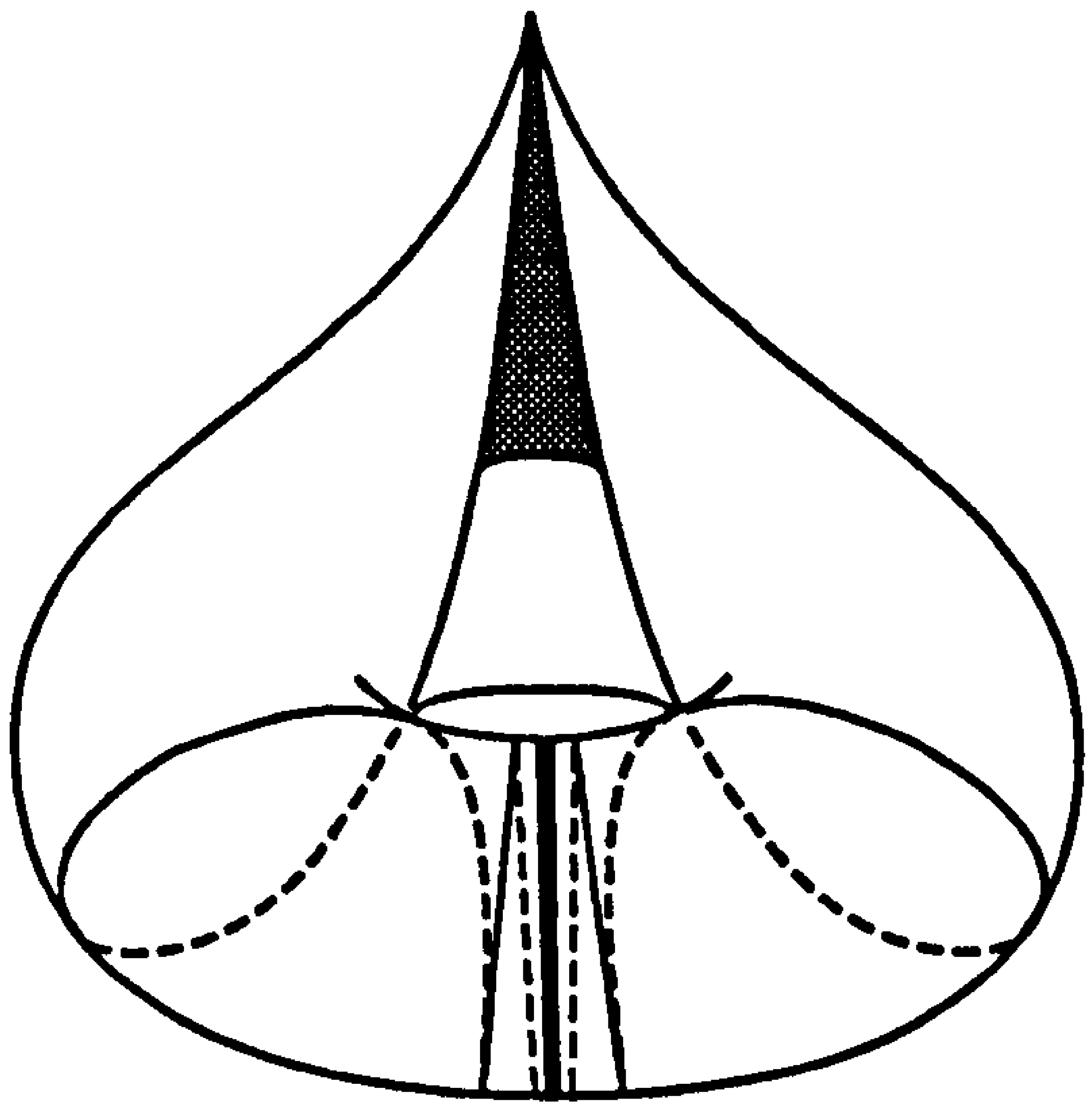

Figure 2.4. A gap corresponding to a certain closed geodesic. The heavy black curve in the lower part of the picture is the closed simple geodesic α. The two geodesics marked each have an end up the cusp and spiral to the closed geodesic from opposite sides. These are the closest geodesics in G_{cusp} to one end of β, the simple geodesic with both ends up the the cusp, which is disjoint from α.

2.3. Approximating Generic Ends

A geodesic in G_{cusp} is contained in an entire simple geodesic. We think of geodesics in G_{cusp} as falling into three classes depending on the behaviour of the other end of this complete geodesic. The other end may:

(a) go up the cusp;

(b) spiral into a closed geodesic;

(c) spiral into a minimal lamination which is not a closed geodesic.

This classification is analogous to that of lemma 1.1.2; however, here we make the distinction between closed geodesics and other compact minimal laminations.

We shall show, at least for the punctured torus, how these three classes arise when we consider how a fixed end is approximated by other points in G_{cusp}.

There are only countably many simple geodesics with both ends up the cusp on a closed punctured surface with finite area. There are also only countably many closed geodesics on a surface of finite area. However, if a surface has an infinite number of closed simple geodesics there must be uncountably many minimal laminations. One way to see this is to argue as follows. Since the surface is of finite type some pair of simple closed geodesics must intersect. From this pair of geodesics we can build a train track with at least two switches and, invoking the theory of train tracks, we can construct uncountably many minimal laminations. For this reason we think of geodesics, whose other end spirals to a minimal lamination which is not a closed geodesic, as being generic in G_{cusp}.

The technique of approximation is a modification of methods found in [CEG].

Diophantine Approximation and the Classification

It is well known that there is a connection between continued fractions and diophantine approximation and the geodesics on the flat torus; see for instance [HW]. In fact there is also an important relationship between the *Markoff spectrum* which arises in diophantine approximation of quadratic irrationals and geodesics on the once punctured torus. This is explained most fully in [Haa] but a useful short account is given in [Ser]. The following shows how our division of G_{cusp} into three classes fits into this classical theory.

Let A be the map $z \mapsto (2z+1)/(z+1)$ and B be the map $z \mapsto (2z-1)/(-z+1)$. These maps generate a free group of rank 2 which we identify with a certain subgroup of $PSL(2,Z)$. The commutator $ABA^{-1}B^{-1}$, $z \mapsto z+6$, is a parabolic element fixing infinity in the upper half space. We think of this group as a discrete group acting on $\mathbf{H}$ with its hyperbolic metric. The quotient surface is a once punctured torus.

Let μ be a vertical line in the upper half plane. The projection of μ to the quotient surface has at least one end up the cusp because infinity is a parabolic point for the group. If the line μ projects to a simple geodesic on the surface then the endpoint on the real axis has a continued fraction expansion of a very special form; the expansion satisfies the so-called *Dixon rules* [Ser]. The set of irrationals which are the endpoints of simple geodesics are called the *Markoff irrationalities*. It can be shown that the Markoff irrationalities are a closed nowhere dense subset of the reals whose Hausdorff dimension is zero [BS].

In addition to detecting when μ projects to a simple curve we can use the continued fraction expansion to say which of the three classes, a,b or c, this projection lies in:

(a) If μ projects to a geodesic with both ends up the cusp in the quotient surface then the endpoint must be at a rational; that is the continued fraction expansion terminates. This is because the orbit of infinity under this group is contained in the rational numbers.

(b) If μ projects to a geodesic which spirals into a closed geodesic on the surface then the finite endpoint of μ is a quadratic irrational; that is the continued fraction expansion is periodic. This is because μ is asymptotic to an axis of a hyperbolic transformation $z \mapsto (az+b)/(cz+d)$ with $a, b, c, d \in Z$. The endpoint of the axis is a fixed point of the transformation and so satisfies a quadratic over Z.

(c) If μ projects to a geodesic which spirals to a lamination which is not a closed geodesic then the endpoint has an infinite aperiodic continued fraction which, nonetheless, satisfies the Dixon rules.

Fixing Notation

When we work in the universal cover of a surface we will identify the cover with the upper half plane. Throughout this section γ will be a simple geodesic with a single end up the cusp. By $\gamma_{\min}$ we mean the minimal lamination that γ spirals into. It will be a running hypothesis that $\gamma_{\min}$ is not a closed geodesic.

We give the surface an orientation. The geodesic γ has a single end up the cusp. For a point on γ we introduce a notion of *left and right* by insisting that the observer face along γ looking towards this end.

With this notion we define the leaf of $\gamma_{\min}$ nearest γ on the right as follows. Since the lamination $\gamma_{\min}$ is not a closed geodesic, the surface obtained by cutting along it has no boundary component which is a circle. The boundary of the cut surface consists of a number of components which are all doubly infinite geodesics. There are exactly two such components in the case of a punctured torus. These components bound portions of the surface called *crowns*. Each crown consist of a number of *spikes*, which are portions of the surface isometric to $\{z : 0 \leq \operatorname{Re} z \leq 1, \operatorname{Im} z \geq p\}$, for some $p > 0$, with the geodesic sides contained in the boundary of the cut surface. For an example of such a cut surface with spikes see figure 1.1. Since γ spirals to $\gamma_{\min}$, the image of γ on the cut surface goes up a spike. For an observer on γ, who is on the image of this spike on the closed surface, *the leaf of $\gamma_{\min}$ nearest γ on the right* is the boundary leaf of $\gamma_{\min}$ on his right as he looks along γ towards the cusp and away from the spike. (Further details of spikes and crowns are in [CB]).

Construction of β and $\beta^{\perp}$

Let β be the leaf of $\gamma_{\min}$ nearest γ on the right. Note that β is necessarily a boundary leaf of $\gamma_{\min}$.

We pick a curve J on the surface (see figure 2.5):

(a) which is simple

(b) which abuts on β and γ such that a portion of β, a portion of γ and J bound an embedded triangle in the surface.

Given a point x on β, we construct a curve as follows: we follow γ from the cusp to $J \cap \gamma$, then we follow J to $J \cap \beta$ and then we follow β to the point x.

We define the curve $\mathbf{C}_x$ to be the geodesic we get when this curve is straightened.

Lemma 2.3.1. *There is a unique x on β such that C_x is perpendicular to β.*

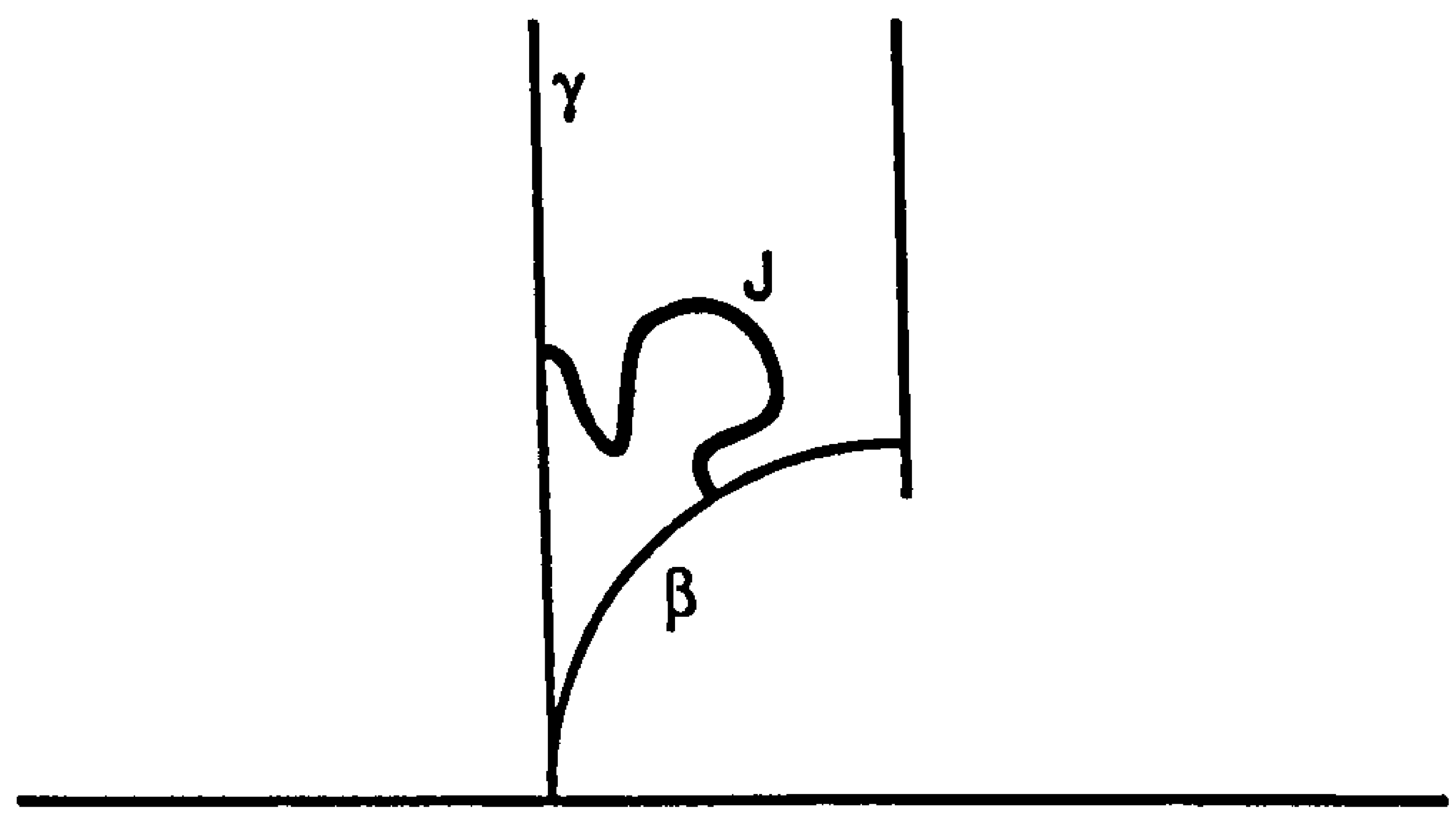

Figure 2.5. A choice for the curve J. The diagram shows a lift of a patch of the surface to the upper half plane with a choice for the curve J indicated. The embedded triangle bounded by a portion of β, a portion of γ and J forms part of this patch.

Lemma 2.3.2. *For each x on β the curve C_x is simple, disjoint from γ and meets β in just one point.*

There is an embedded triangle in the surface bounded by γ, C_x and the portion of β asymptotic to γ and with endpoint the intersection of β and C_x.

Proof of 2.3.2: We cut along both γ and $\gamma_{\min}$ to obtain a surface of finite type. This cut surface is convex and has geodesic boundary.

The image of J is still simple but now it abuts on two boundary geodesics, one corresponding to γ and the other to β. We can thus construct a connected simple curve on the cut surface whose pre-image on the original surface is in the same proper homotopy class as C_x. Since the cut surface is convex we can straighten this curve to a geodesic. The pre-image of this geodesic is just C_x.

This geodesic has one endpoint at the ideal vertex corresponding to the cusp and the other endpoint on the boundary. Since there is a simple curve in its proper homotopy class, it is simple. So C_x is simple.

Also since the boundary is geodesic and has only ideal vertices, this geodesic can only intersect it once. This means C_x intersects β and does not intersect γ. [2.3.2]

Definition 2.3.3 (perpendicular to β in the homotopy class of γ). We denote by $\beta^\perp$ the unique geodesic which intersects β perpendicularly in a single point, and which, together with γ and some portion of β, bounds an embedded triangle. We say $\beta^\perp$ is *the geodesic perpendicular to β in the homotopy class of γ.*

Remark: We note that there is a simple geodesic strictly containing $\beta^\perp$.

Making Simple Curves

Definition 2.3.4. On the surface M let $\beta^\perp(\epsilon)$ be the geodesic containing $\beta^\perp$ such that the length of the curve $\beta^\perp(\epsilon)\backslash\beta^\perp$ is ϵ.

Henceforth we fix a small positive value of ϵ so that $\beta^\perp(\epsilon)$ and so that the angle made when γ intersects $\beta^\perp(\epsilon)$ is sufficiently close to a right angle.

Lemma 2.3.5 (infinite intersections). *Let $T > 0$. The intersection of the curve $\{\gamma(t) : t > T\}$ with $\beta^{\perp}(\epsilon)$ is an infinite set.*

Proof of 2.3.5: If $\{\gamma(t) : t > T\}$ and $\beta^{\perp}(\epsilon)$ are disjoint then there is an open ball centre $\beta^{\perp} \cap \beta$ disjoint from $\{\gamma(t) : t > T\}$. (This is because γ and β are disjoint.) But this contradicts the fact that β is contained in the closure of γ. 2.3.5

Definition 2.3.6 (height, closest approach). Let x be a point on $\beta^{\perp}(\epsilon)\backslash\beta^{\perp}$. We define the *height* of x, $\Theta(x)$, as the distance from x to $\beta^{\perp} \cap \beta$. Since ϵ is small, the distance can be measured along $\beta^{\perp}(\epsilon)$.

A *closest approach for γ to the point $\beta^{\perp} \cap \beta$* is a value of the parameter t, T say, such that $\gamma(T)$ lies on $\beta^{\perp}(\epsilon)$ and such that $\Theta(\gamma(T))$ is less than $\Theta(\gamma(t))$ for all $t < T$, such that $\gamma(t) \in \beta^{\perp}(\epsilon)$.

Lemma 2.3.7 (closest approach gives simple). *Let T be a closest approach for γ. The union of $\{\gamma(t) : t < T\}$ and the portion of $\beta^{\perp}(\epsilon)$ which connects $\gamma(T)$ to the cusp is a simple curve.*

Proof of 2.3.7: The curves $\{\gamma(t) : t < T\}$ and the portion of $\beta^{\perp}(\epsilon)$ are simple because they are contained in the simple curves, γ and $\beta^{\perp}(\epsilon)$ respectively. These two curves meet only in the single point $\gamma(T)$ because γ is disjoint from $\beta^{\perp}$ and T is a closest approach. 2.3.7

Definition 2.3.8 (join, double join). We call the union of $\{\gamma(t) : t < T\}$ and the portion of $\beta^{\perp}(\epsilon)$ which connects $\gamma(T)$ to the cusp the *join*. Let S and T be closest approaches for γ. Let α be the curve obtained by following $\beta^{\perp}(\epsilon)$ from the cusp to $\gamma(S)$ then following γ to $\gamma(T)$ then going up the cusp along $\beta^{\perp}(\epsilon)$. We say the curve α is the *double join between S and T*.

Lemma 2.3.9 (double join gives simple). *Let S and T be consecutive closest approaches for γ. The geodesic in the proper homotopy class of the double join between S and T is simple.*

Proof of 2.3.9: Since it is easy to see that the double join is a simple curve we need only prove that there is a geodesic in its proper homotopy class. To establish this we must show that some lift of this double join has distinct endpoints in **H**; this can be done as follows. As $\beta^{\perp}$ meets β perpendicularly and ϵ was chosen suitably small, the angles A,B are approximately right angles. So this lift has distinct endpoints (see figure 2.6).

2.3.9

Definition 2.3.10 (approach from the left). We give the ambient surface, M, an orientation. The geodesic $\beta^{\perp}(\epsilon)$ has a single end and this end goes up the cusp. Let α be an infinite geodesic with a distinguished end. Suppose α meets $\beta^{\perp}(\epsilon)$ at the point x. The intersection point x is *an approach from the left* if the the end of α is on the left side of an observer at the point x who faces along $\beta^{\perp}(\epsilon)$ looking towards the cusp.

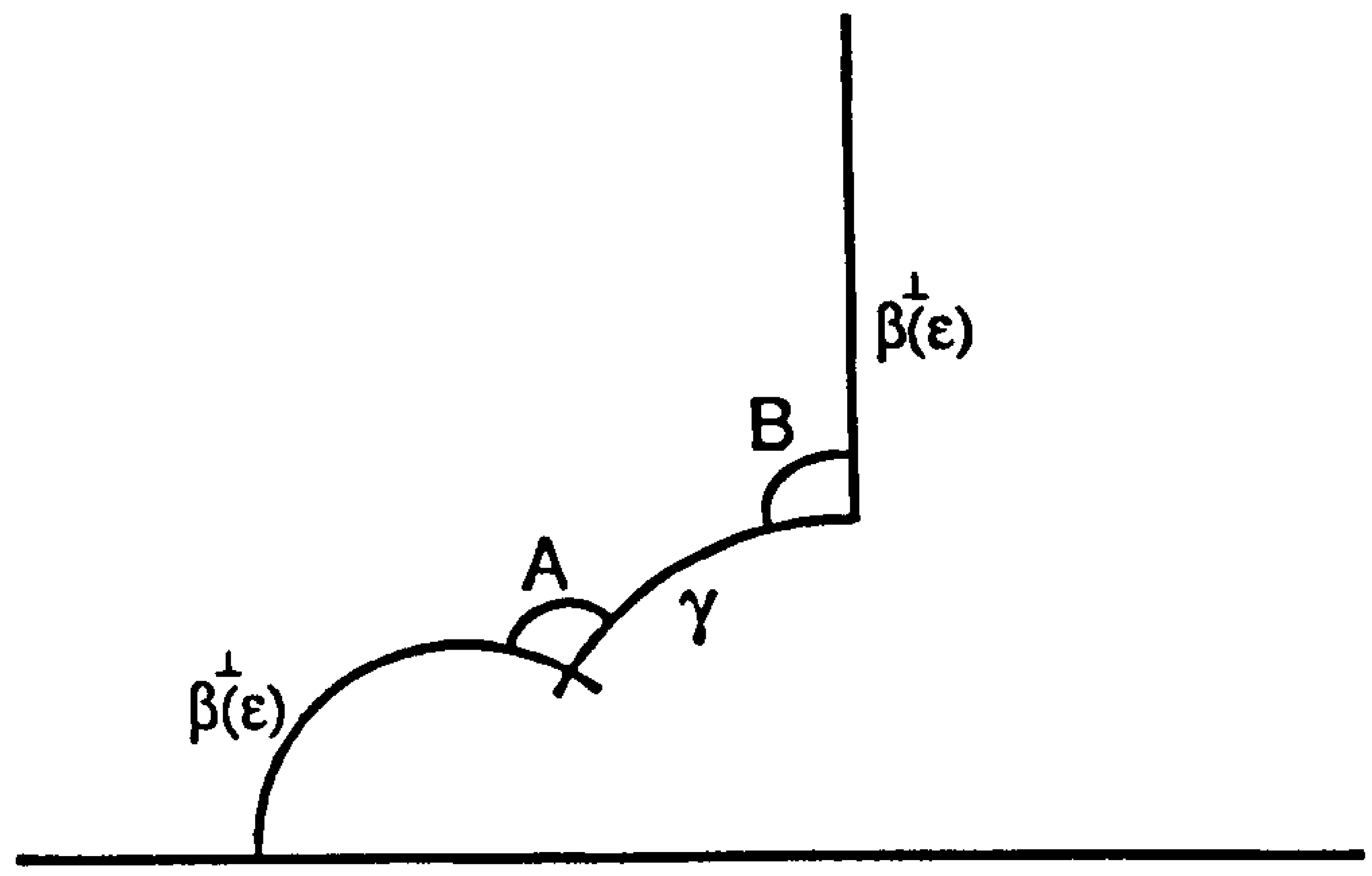

Figure 2.6. A lift of the double join. The figure shows a lift of the double join to H. The marked angles, A and B, are approximately right angles, consequently the lift has distinct endpoints on the ideal boundary.

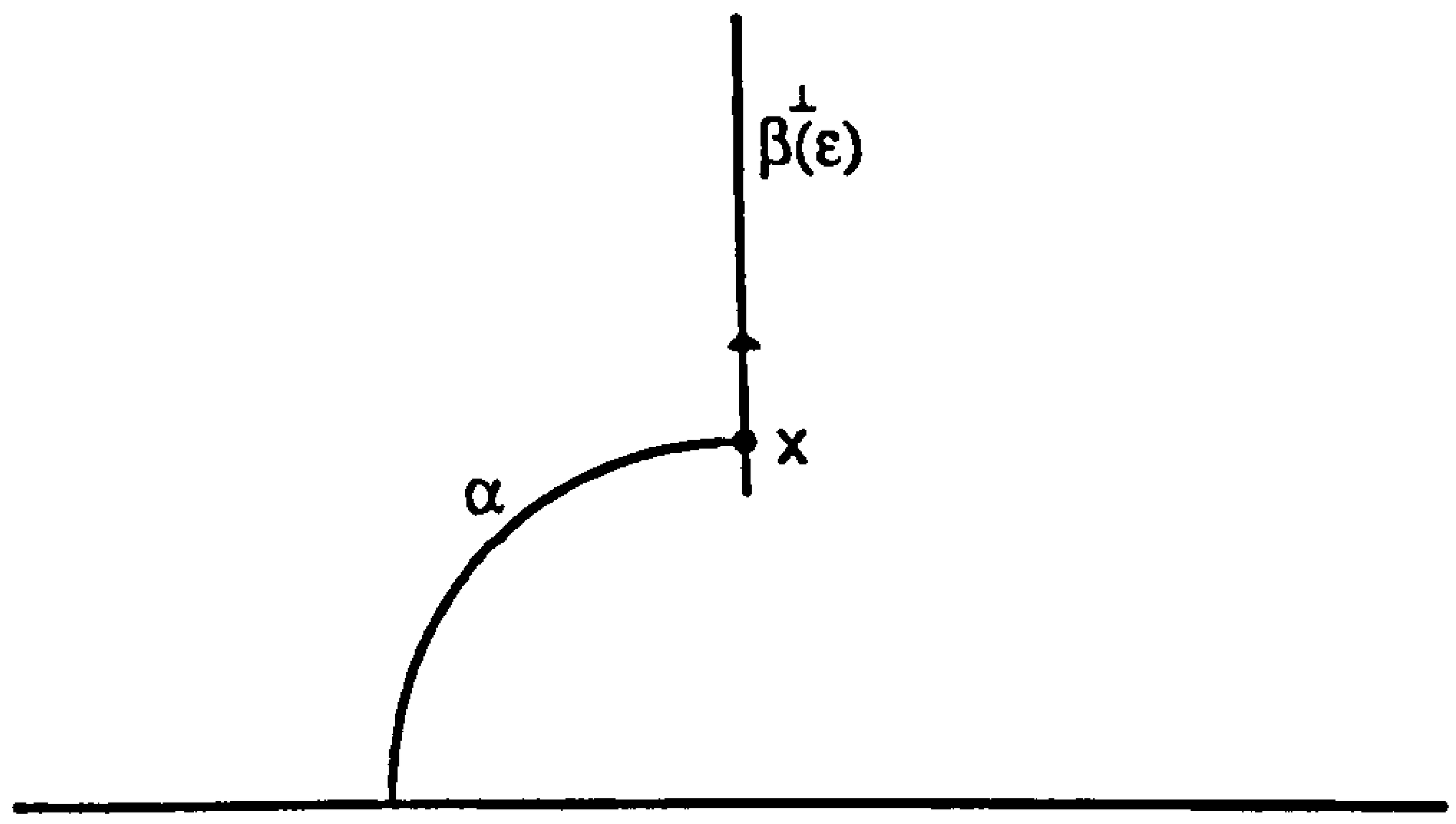

Figure 2.7. An approach from the left. An approach from the left by the curve α. The arrow represents the direction the observer at x is looking in.

Lemma 2.3.11 (approximation estimates). *Let α be a geodesic with a distinguished end which does not intersect β or $\beta^{\perp}$ and has an endpoint x on $\beta^{\perp}(\epsilon)$. Suppose α is an approach from the left.*

The curve which is the union of α from its ideal endpoint (on the distinguished end) to x, and the portion of $\beta^{\perp}(\epsilon)$ between x and the cusp, can be straightened to a geodesic with at least one end up the cusp. There is a constant, K, depending only on β and γ, such that one end of this geodesic is closer than $K\Theta(x)$ to γ.

Proof of 2.3.11: We lift to H so that the horocycle of length 1 is contained in $\{z : \operatorname{Im} z = 1\}$ and γ lifts to the line $\hat{\gamma} = \{z : \operatorname{Re} z = 0\}$. There is a unique lift of β, $\hat{\beta}$, which is a semicircle with endpoints 0 and p , for some $0 < p \leq 1/2$. There is a lift, of $\beta^{\perp}(\epsilon)$ contained in the line $\{z : \operatorname{Re} z = p/2\}$.

Under our hypothesis there is a lift of α lying entirely below the semicircle $\hat{\beta}$, which intersects $\beta^{\perp}(\epsilon)$.

We note that 0 as an endpoint of $\hat{\gamma}$. So the distance, as measured by the metric on ends going up the cusp, between the geodesic γ and the end of the join of α and $\beta^{\perp}(\epsilon)$ is the distance between 0 and the endpoint of $\hat{\alpha}$ on $\mathbf{R}$. Since $\hat{\alpha}$ lies entirely below $\hat{\beta}$ this distance is less than the Euclidean diameter of $\hat{\beta}$ minus the Euclidean diameter of $\hat{\alpha}$.

The Euclidean diameter of $\hat{\alpha}$ is greater than twice the imaginary part of any point on it. In particular it is greater than $2\operatorname{Im}\hat{x}$, where $\hat{x}$ is the intersection x of $\beta^{\perp}(\epsilon)$ and $\hat{\alpha}$. The point $\hat{x}$ projects to the intersection of α and $\beta^{\perp}(\epsilon)$ on the surface. We calculate $\Theta(x)$ explicitly:

$$\begin{aligned}\Theta(x) &= \int_{\operatorname{Im}\hat{x}}^{p/2} dy/y \\ &= \log\frac{p}{2} - \log\operatorname{Im}\hat{x}.\end{aligned}$$

So

$$\operatorname{Im}\hat{x} = \frac{p}{2}e^{-\Theta(x)}.$$

Thus

$$\begin{aligned}\operatorname{diameter}\hat{\beta} - \operatorname{diameter}\hat{\alpha} &\leq p(1 - e^{-\Theta(x)}) \\ &\leq p\Theta(x).\end{aligned}$$

It follows that we can take $K = p$. We remark that $-\log(p/2)$ is just the length of the portion of $\beta^{\perp}$ outside the horocycle of length 1, so K depends only on $\beta^{\perp}$. 2.3.11

Lemma 2.3.12 (length estimate). *Let $\epsilon > 0$. Let γ be the geodesic $\{z : Im z = 0\}$ and β be the geodesic with endpoints 0 and 1. Let z be the highest point on β, x the point at the same Euclidean height as z on γ, and y any point on γ.*

If x' and y' are points of $\boldsymbol{H}$ such that $d(z, x') < \epsilon$ and $d(y, y') < \epsilon$ then

$$d(x', y') \geq d(x, y) - 2\epsilon - 1,$$

where d is the hyperbolic metric on $\boldsymbol{H}$.

Proof of 2.3.12: The points z and x are joined by a horocyclic curve of length $\frac{1}{2}$ so $d(z, x) < 1$. The lemma then follows from an application of the triangle inequality. 2.3.12

We want to show that there are sequences, r_n and l_n in G_{cusp}, which approximate γ from the right and the left respectively. In the proof of the following theorem a method is given for constructing a sequence r_n. This method can easily be adapted to show the existence of a sequence l_n.

Theorem 2.3.13 (approximating from right). *Let γ be a simple geodesic with a single end up the cusp and its other end spiralling into a minimal lamination which is not a closed geodesic. Then the geodesic γ can be approximated from the right by simple geodesics with both ends up the cusp.*

Proof of 2.3.13: The geodesic γ has a single end up the cusp; we take this as its distinguished end in the manner of definition 2.3.10. Let $T_1 < T_2 < T_3 < \ldots$ be a complete list of the closest approaches of γ along $\beta^{\perp}(\epsilon)$.

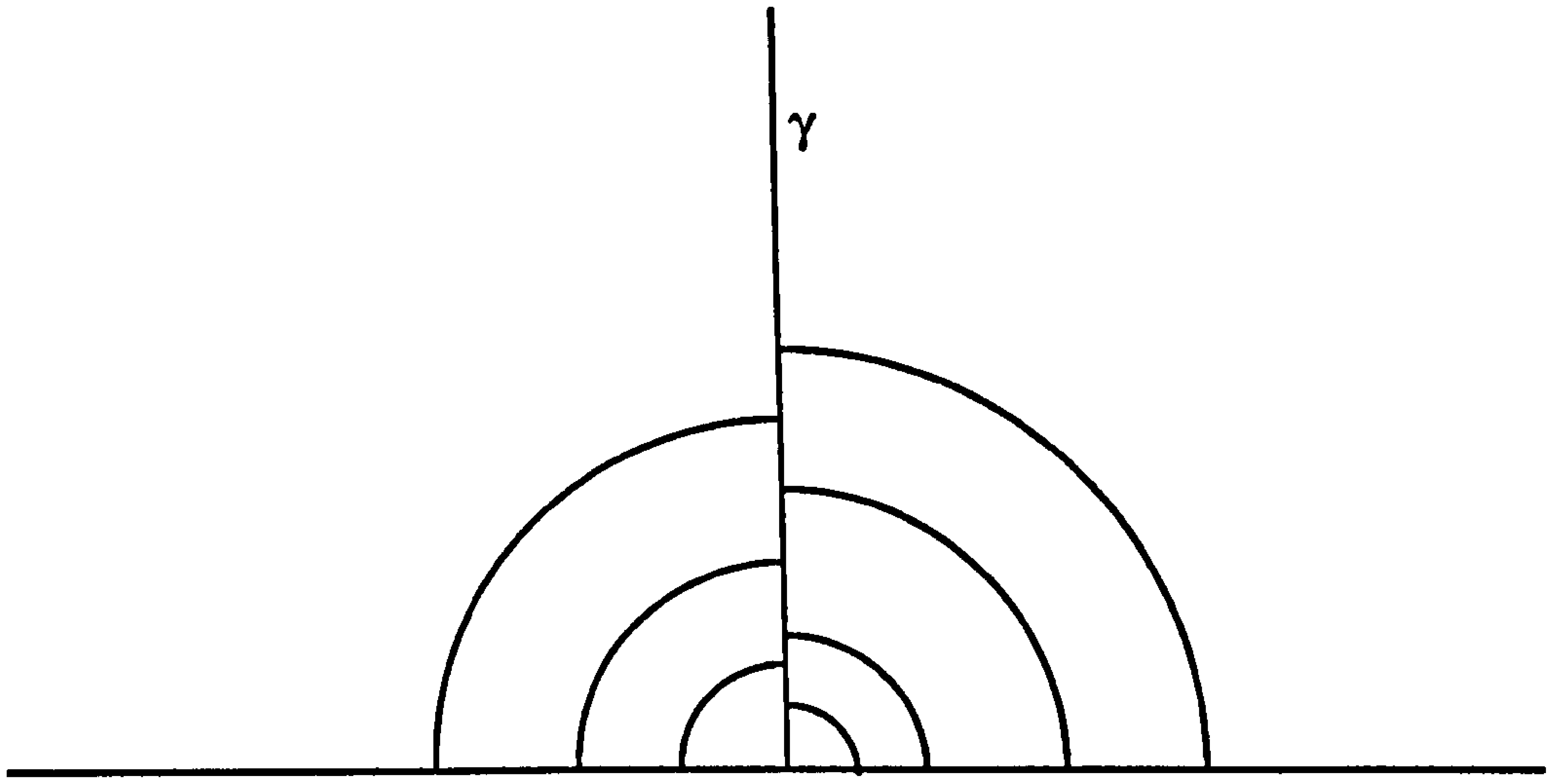

Figure 2.8. Approaches from left give approximations from right. The vertical line is a lift of the geodesic γ and the circular arcs are lifts of $\beta^{\perp}(\epsilon)$ corresponding to a sequence of closest approaches. The subsequence of lifts of $\beta^{\perp}(\epsilon)$ which have endpoints on the right of the lift of γ in fact meet γ in approaches from the left; from these we can construct a sequence of closed simple geodesics approximating γ from the right by performing the join construction of the last section.

If there is an infinite subsequence of closest approaches, T_{i_n}, which are from the left of $\beta^{\perp}(\epsilon)$ then there is a corresponding sequence of geodesics with both ends up the cusp approaching γ from the right. Such a sequence is obtained by following γ from the cusp to $\gamma(T_{i_n})$ then going back up the cusp along $\beta^{\perp}(\epsilon)$ (see figure 2.8).

Suppose,for a contradiction, that there exists an N such that for all $i \geq N$, T_i is a closest approach from the right.

There is an embedded geodesic triangle in the surface which is bounded by γ, $\beta^{\perp}$ and a portion of β. This triangle has the cusp as an ideal vertex. We foliate the triangle with leaves which are horocyclic based at the cusp point. Let x be the point of γ on the same leaf as $\beta^{\perp} \cap \beta$ and let $t(x)$ be the parameter value corresponding to x.

We now choose N (bigger if necessary) so that T_N is greater than $t(x) + 2\epsilon + 1$. Let $n > N$.

We lift to $\mathbf{H}$ (figure 2.9) so that γ is $\{z : \operatorname{Re} z = 0\}$. Conjugate the covering group if necessary so that there is a lift of β with endpoints $0, 1$. There is then a lift $\beta^{\perp}(\epsilon)$ contained in $\{z : \operatorname{Re} z = \frac{1}{2}\}$. We lift $\gamma(T_n)$ to a point on this line. This lift of $\gamma(T_n)$ lies on a unique lift of γ (different from $\{z : \operatorname{Re} z = 0\}$); we denote this by γ'. There is a lift of $\gamma(T_{n+1})$, which we will also denote by $\gamma(T_{n+1})$, which is a point on γ'.

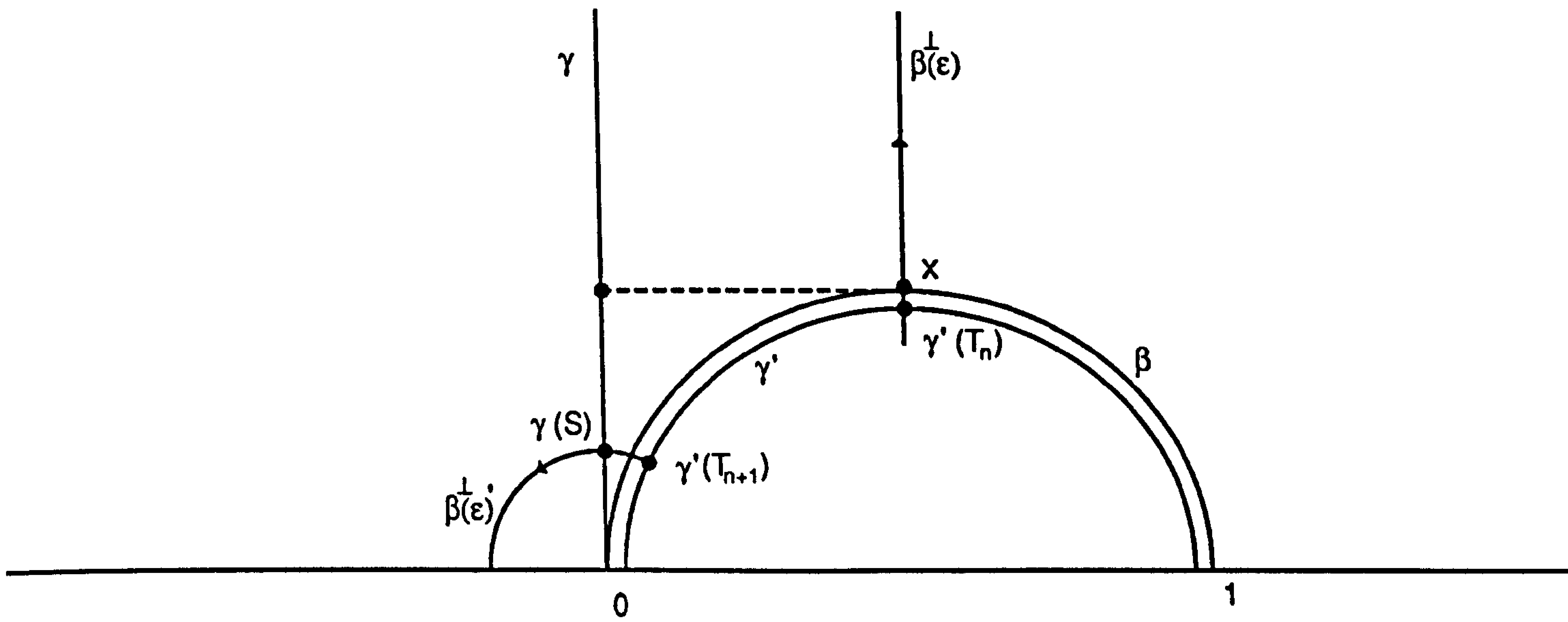

Figure 2.9. Diagram of lifts. This is a diagram of the lifts of the curves used in the proof. The cusp point has been lifted to the point at infinity. There are two lifts of γ; the first, labelled γ, is a vertical line and the second , γ', is the smaller of the two semicircles. There are also two lifts of $\beta^{\perp}(\epsilon)$; one is a vertical line segment labelled $\beta^{\perp}(\epsilon)$ and the other, $\beta^{\perp}(\epsilon)'$, is a circular arc. The dotted curve is a horocyclic segment of length 1.

Having chosen this lift $\gamma(T_{n+1})$ there is a lift of $\beta^{\perp}(\epsilon)$, $\beta^{\perp}(\epsilon)'$, through this point. Since T_{n+1} is an approach from the right to $\beta^{\perp}(\epsilon)$ this lift has its ideal endpoint to the left of both endpoints of γ', If this endpoint were negative (that is if the double join between T_n and T_{n+1} were to approximate γ from the left) then the intersection of $\beta^{\perp}(\epsilon)'$ and γ would give a point on the surface on γ which was a closer approach than T_{n+1}. We show that this cannot happen by proving that the parameter value along γ of this intersection would be less than T_{n+1}.

Let S be the parameter value of this intersection. Since γ never intersects $\beta^{\perp}$,

$$d(\gamma(T_{n+1}), \gamma(S)) < \Theta(T_{n+1}) < \epsilon$$

By lemma 2.3.12,

$$T_{n+1} - T_n > S - t(x) - 2\epsilon - 1.$$

So, since $n > N$,

$$\begin{aligned} T_{n+1} &> S + T_n - t(x) - 2\epsilon - 1 \\ &> S + T_n - T_N \\ &> S. \end{aligned}$$

This contradicts the hypothesis that T_{n+1} was a closest approach. So the join between T_n and T_{n+1} approximates γ from the right.

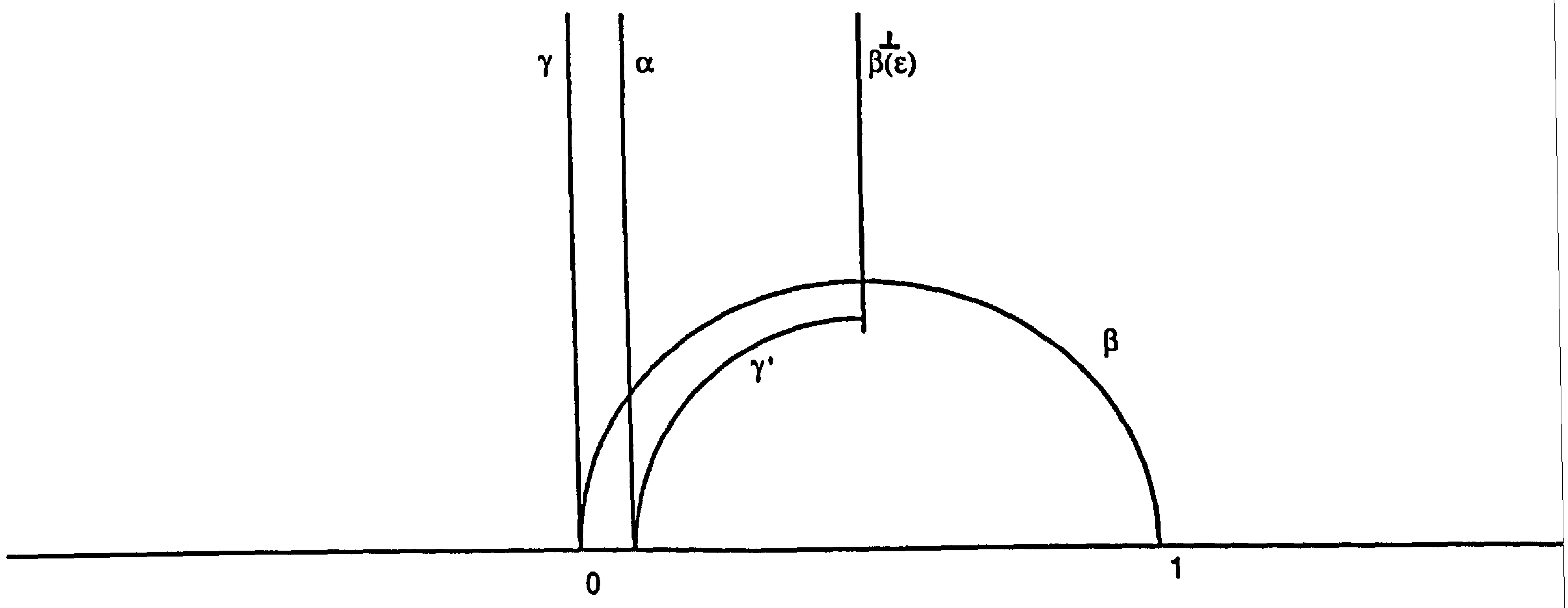

Figure 2.10. The lift of the curve α. The diagram shows a lift of the curve α. This curve is obtained by joining a portion of $\beta^{\perp}(\epsilon)$ to γ'. There is a lift of the double join which is a vertical line with its endpoint sandwiched between 0 and the endpoint of α.

Let α be the geodesic we get by straightening the curve obtained by following $\beta^{\perp}(\epsilon)$ from the cusp to $\gamma(T_n)$ then going along the end of γ' which is on our left as we look back towards the cusp along $\beta^{\perp}(\epsilon)$. It follows from lemma 2.3.11 that this geodesic is closer than $K\Theta(\gamma(T_n))$ to γ. The hypothesis that γ always approaches $\beta^{\perp}(\epsilon)$ from the left forces the endpoint of the double join between T_n and T_{n+1} to be on the left of the endpoint of α. The double join is a better approximation to γ than this curve α.

2.3.13

2.4. Approximating Ends by Dehn Twists

We use *Dehn twists* to show that an end contained in a simple geodesic whose other end spirals to a closed simple geodesic, is not isolated in G_{cusp}. If the closed geodesic is separating then the sequence we construct consists of geodesics each with exactly one end up the cusp. If the closed geodesic is non-separating then each geodesic in the sequence has both ends up the cusp. In either case each geodesic in the sequence meets the closed geodesic exactly once.

The necessary background material relating to Dehn twists can be found in [CB]. More detailed accounts are given in [Ker] or [Goo].

Lemma 2.4.1 (curve that cuts only once). *Let M be a once punctured surface of finite volume and without boundary. Let α be a simple geodesic with* *one end up the cusp and the other end spiralling to a closed geodesic γ. There is a complete simple geodesic, β, such that*

(a) it meets γ in a single point;

(b) for any lift of α to H there is a lift, γ', of γ and a lift, β', of β such that γ' is asymptotic to the end of the lift of α which projects to an end spiralling into γ; β' is asymptotic to the end of α which projects to an end lying in the cusp region; γ' and β' intersect.

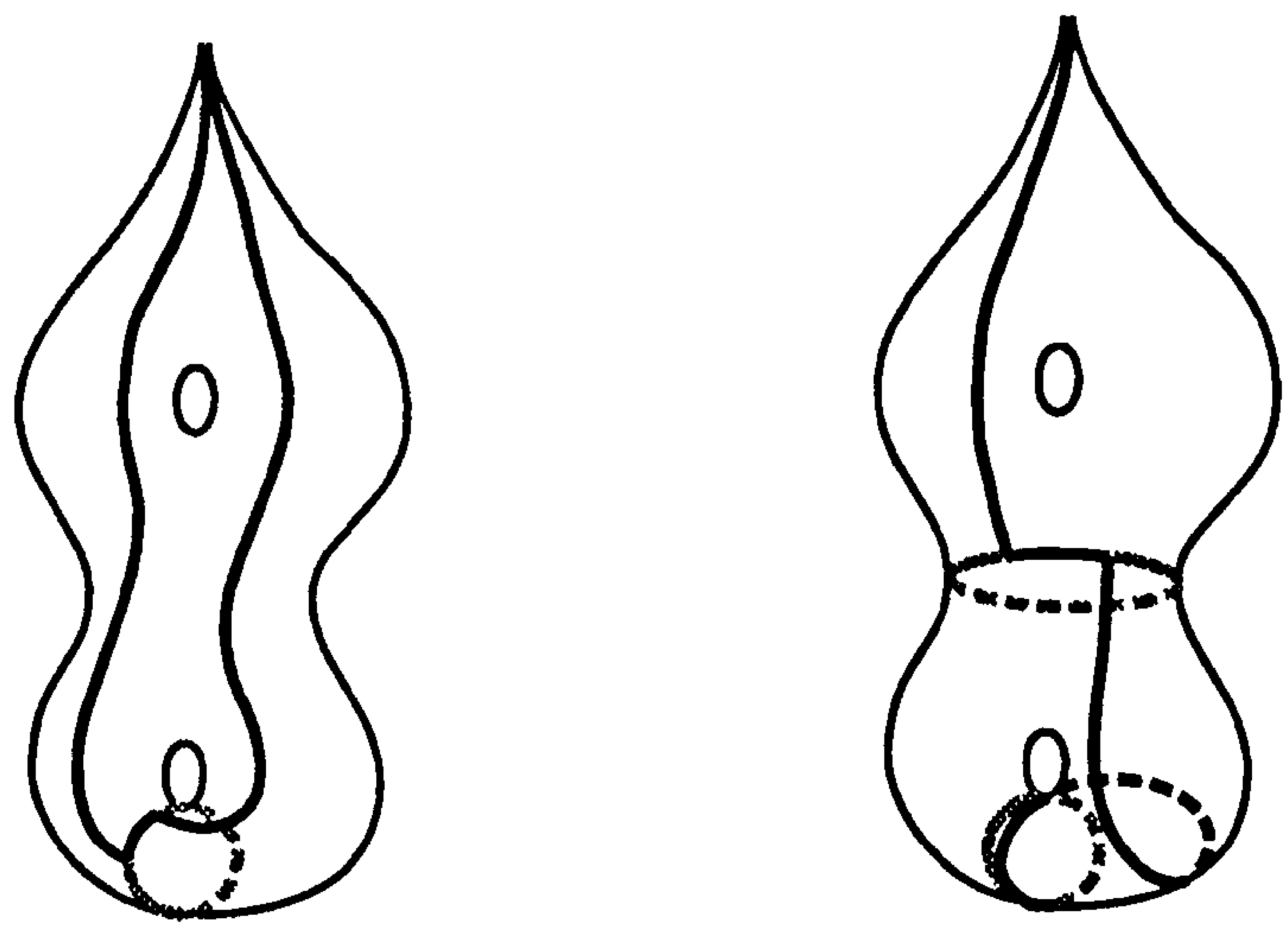

Figure 2.11. The two different cases. The figures represent the same once punctured surface of genus 2 but with different choices for the geodesic α. In the first case α spirals to a non-separating geodesic and in the second α spirals to a separating geodesic. The broken curves are the curves β of the lemma.

Proof of 2.4.1: Let ϵ be the injectivity radius of the thick part of the surface. We cut along γ.

Suppose first that γ is non-separating; that is, the cut surface is connected. Let C be a collar round γ of width ϵ. The curve α must eventually enter this collar, so there is a point of α on the boundary of C. Let I be the shortest path from this point to γ; this path is necessarily simple and disjoint from the portion of α exterior to the collar. Let β_1 be the curve which is the union of I and $\alpha\backslash C$. Choose β_2 to be any simple geodesic disjoint from β_1 with an end up the cusp and one endpoint on the other boundary component of the cut surface. The image of β_1 and β_2 on the surface is a pair of disjoint curves each with an endpoint on γ. Take the curve β to be the union of β_1, β_2 and a segment of γ joining the two endpoints. When we straighten β we get a geodesic satisfying the conditions of the lemma.

Suppose that γ is separating. Construct the curve β_1 as before. The component of the cut surface which does not contain the cusp has at least one closed geodesic. There is a simple geodesic β_2 with an endpoint on γ and a single end spiralling into some closed geodesic in this component. The images of β_1, β_2 on the surface are a pair of disjoint curves each with an endpoint on γ, as in the previous case. The curve β is the union of β_1, β_2 and a segment of γ joining the two endpoints. When we straighten β we get a geodesic satisfying the conditions of the lemma. [2.4.1]

Definition 2.4.2 (Dehn twist). Let γ be a closed simple geodesic on a surface. There is a regular neighborhood of γ which is homeomorphic to an annulus. Parameterize this neighborhood as $\{[r,\theta] : 1 \leq r \leq 2\}$. The *Dehn twist* round γ is the homeomorphism which is the identity off this annulus and the map $[r,\theta] \mapsto [r, \theta + 2\pi r]$ on the annulus.

Remarks: The image of a geodesic β which intersects γ, under a Dehn twist round γ, is not a geodesic. However, it is a piecewise smooth curve and there is a unique geodesic determined by straightening this. If the geodesic β is simple then the geodesic determined by its image is simple. If the geodesic β intersects γ exactly n times then the geodesic determined by the image of β under a Dehn twist round γ intersects γ exactly n times

Theorem 2.4.3 (approximating by twisting). *Let M be a closed punctured surface. Let α be a simple geodesic with one end up the cusp and the other end spiralling to a closed geodesic γ. Then there is a sequence of points of G_{cusp} converging to α.*

These points can be taken to be ends of a sequence of geodesics which are determined by the images of a certain geodesic under iterated Dehn twists round γ. The choice of this geodesic depends on α.

Proof of 2.4.3: We lift to $\mathbf{H}$. Conjugate the group of covering transformations for M so that the line $\hat{\alpha} = \{z : \operatorname{Re} z = 0\}$ is a lift of α, and the semi-circle, $\hat{\gamma}$, with endpoints 0 and 1 is a lift of γ. Let A be the hyperbolic transformation which covers this lift. We may suppose that the attracting fixed point of A is 0, that is, it is the same as the endpoint of $\hat{\alpha}$ on the real axis. Let β be the geodesic constructed in lemma 2.4.1. There is a lift of β, $\hat{\beta}$, which is a vertical line intersecting $\hat{\gamma}$.

The Dehn twist round γ induces an earthquake on $\mathbf{H}$; the quake locus is the lamination of $\mathbf{H}$ consisting of all lifts of γ. Let $\mathrm{T}_\gamma\mu$ denote the geodesic determined by straightening the image under this earthquake of μ.

Since β intersects γ only once on the surface $\hat{\beta}$ intersects exactly one leaf of the lamination of $\mathbf{H}$ by lifts of γ, Because of this it is easy to calculate the image of $\hat{\beta}$ under the earthquake of $\mathbf{H}$ induced by the Dehn twist as follows. If $\hat{\beta}$ has finite endpoint x then $\mathrm{T}_\gamma\hat{\beta}$ is the vertical line ending at Ax.

The portion of $\mathrm{T}_\gamma\hat{\beta}$ lying in $\{z : \operatorname{Im} z > 1\}$ is nearer to $\hat{\alpha}$ (the lift of α) than the portion of $\hat{\beta}$ which also lies in $\{z : \operatorname{Im} z > 1\}$.

The projection of $\mathrm{T}_\gamma\hat{\beta}$ to the surface is a simple geodesic which intersects γ just once. So we may repeat this twisting procedure to obtain a sequence, $\mathrm{T}^n_\gamma\hat{\beta}$ converging to $\hat{\alpha}$. Let $cl(\mathrm{T}^n_\gamma\hat{\beta})$ be the closure of $\mathrm{T}^n_\gamma\hat{\beta}$; each $cl(\mathrm{T}^n_\gamma\hat{\beta})$ is a lamination of $\mathbf{H}$. The sequence of the laminations on M covered by the $cl(\mathrm{T}^n_\gamma\hat{\beta})$ converges to a lamination containing α. [2.4.3]

2.5. Gaps

Definition 2.5.1 (G_{one}, G_{two}, gaps in cusp region). Let G_{one} be the set of ends of all complete simple geodesics each with a single end up the cusp and let G_{two} be the set of all ends of complete simple geodesics with both ends up the cusp. A *gap* is a maximal strip in the complement of G_{one} in the cusp region.

Theorem 2.5.2 (geodesic which bounds a gap). *A geodesic arc which bounds a gap is contained in a complete simple geodesic which spirals into a closed geodesic.*

Proof of 2.5.2: Let γ be a geodesic which bounds a gap. Since the set of geodesic laminations is closed in the Hausdorff topology the complete geodesic that contains γ must be simple. Suppose γ spirals into a lamination which is not a closed geodesic. We apply theorem 2.3.13 to get a sequence which converges to γ from the right and as was indicated in the text preceding theorem 2.3.13 we can construct a sequence which converges from the left in an analogous manner. This contradicts the hypothesis that γ bounds a gap. [2.5.2]

Let M be a once punctured torus. The set G_{cusp} is the union of G_{one} and G_{two}. By theorem 2.2.5 every point of G_{two} is isolated in G_{cusp}. So G_{two} is open and G_{one} is closed in G_{cusp}. Recall that the intersection of G_{cusp}, and therefore G_{one}, with a horocycle is a closed subset of the horocycle. Every closed simple geodesic on M is non-separating so by theorem 2.4.3 and theorem 2.3.13 G_{two} is dense in G_{cusp}.

In the next section we will make use of a theorem of J. Birman and C. Series [BS] which implies that the intersection of G_{one} with a small horocycle has Hausdorff dimension 0. For the moment we prove the following weaker assertion using the results of the preceding sections.

Theorem 2.5.3 (no isolated points). *Let M be a once punctured torus. Then G_{one} (with the induced metric from G_{cusp}) has no isolated points.*

Proof of 2.5.3: Let $\alpha \in G_{\text{one}}$ and μ be the minimal lamination it spirals to. There is a sequence of geodesics in G_{two}, β_n, which converge to a lamination containing α (if μ is a closed geodesic this follows from theorem 2.4.3 otherwise it follows from theorem 2.3.13). The geodesic β_n is disjoint from a unique closed simple geodesic γ_n. For each γ_n there are four geodesics in G_{one} which spiral into it; these bound a pair of gaps which contain the ends of β_n. This means that there is a point of G_{one} nearer to α than β_n is. [2.5.3]

Corollary 2.5.4 (Cantor set). *Let M be a once punctured torus. Then the intersection of G_{one} with a horocycle of length less than 2 is a Cantor set.*

Proof of 2.5.4: Let C be a horocycle of length less than 2 on M. Since G_{one} is closed as a subset of G_{cusp} with its metric, $G_{\text{one}} \cap C$ is closed as a subset of C. By theorem 2.5.3 $G_{\text{one}} \cap C$ has no isolated points.

We see that $G_{\text{one}} \cap C$ contains no interval as follows. Suppose there is a point x interior to some interval contained in $G_{\text{one}} \cap C$. Since G_{two} is dense in G_{cusp} there is a sequence, x_n, converging to x with each x_n in the intersection of G_{two} and C. By theorem 2.2.5 each of the x_n is interior to a gap; this is a contradiction. $\boxed{2.5.4}$

Corollary 2.5.5 (geodesics which bound same gap). *Let J be a gap on any puctured torus. The two geodesics which bound J spiral into the same closed geodesic.*

Proof of 2.5.5: Let γ be one of the geodesics which bound J. By theorem 2.2.5 there is a gap K which is bounded by γ and another geodesic which spirals to the same closed geodesic as γ. If J is not contained in K then γ is an isolated point of G_{one}; this is impossible because of theorem 2.5.3. So J $\subset K$ and, by maximality of J, the two must be equal. $\boxed{2.5.5}$

Remark: Each closed simple geodesic has exactly four geodesics in G_{one} which spiral to it. These four geodesics bound a pair of gaps, so we have now established a 2-1 correspondence between gaps and closed simple geodesics on the torus.

2.6. An Identity for Tori

It is well known that the pointset of a geodesic lamination on a hyperbolic surface has measure zero. In contrast to this a flat torus has a lamination consisting of lines of irrational slope which is full measure. Through each point on a flat torus there are infinitely many geodesic laminations. In [Jor] it is shown that if two geodesics pass through a point x of a hyperbolic surface then infinitely many geodesics must pass through x. So, in some sense, a hyperbolic surface resembles the flat torus. However, the following theorem shows that for almost all points x of a hyperbolic surface it is impossible to find a geodesic lamination passing through x. For the proof see [BS].

Theorem 2.6.1 (Birman-Series). *Let G be the set of all simple geodesics on a hyperbolic surface. The set S of points which lie on a geodesic $\gamma \in G$ has Hausdorff dimension 1.*

Corollary 2.6.2 (gaps are full measure). *The area of a cusp region is equal to the area of the union of of all maximal gaps.*

The following theorem is an immediate consequence of the above and the results of section 2.5.

Theorem 2.6.3 (identity for tori). *Let M be any punctured torus then*

$$\sum_{\gamma} \frac{1}{1+\exp|\gamma|} = \frac{1}{2},$$

where the sum is over all closed simple geodesics γ.

Proof of 2.6.3: We choose a cusp region and compute its area as the sum of areas of gaps as follows.

Every gap corresponds to a unique closed geodesic and every closed geodesic has exactly two gaps associated to it. Each of the gaps associated to the geodesic γ has area $1/(1+\exp|\gamma|)$ times the area of the cusp region . The area of the cusp region is exactly the sum of the areas of the gaps so, by theorem 2.2.5

$$\sum_{\gamma} \frac{2}{1+\exp|\gamma|} = 1.$$

2.6.3

Bibliography

[Abi] W. Abikoff. *The Real analytic theory Of Teichmuller space,Lecture notes in mathematics,820.* Springer, New York, 1980.

[BS] J. Birman and C. M. Series. Geodesics with Bounded Intersection are Sparse. *Topology* **24 no. 2**(1985), 217–225.

[CEG] R. D. Canary, D. B. A. Epstein, and P. Green. Notes on notes of Thurston. In *Analytical and geometric aspects of hyperbolic space, LMS Lecture Notes Series 111*, pages 3–92. Cambridge University Press, Cambridge, 1987.

[CB] A. J. Casson and S. A. Bleiler. *Automorphisms of Surface after Nielsen and Thurston, LMS Student Texts 9.* Cambridge University Press, Cambridge, 1988.

[Goo] O. A. Goodman. Doctoral Thesis. Ph.D. thesis, University of Warwick, 1989.

[Haa] A. Haas. Diophantine approximation on hyperbolic Riemann surfaces. *Acta Mathematica* **156 no. 1-2**(1986), 33–82.

[HW] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers.* Clarendon Press, Oxford, 1962.

[Jor] T. Jorgensen. Closed geodesics on Riemann Surfaces. *Proc. Amer. Math Soc.* **72**(1978), 140–142.

[Ker] S. P. Kerckhoff. The Nielsen realisation problem. *Annals of Math* **117**(1983), 235–265.

[Ser] C. M. Series. The geometry of Markoff numbers. *Mathematical Intelligencer* **7 no. 3**(1985), 20–29.

[Thu] W. P. Thurston. *The Geometry and Topology of 3-Manifolds.* To Appear.