Diophantine approximation on hyperbolic Riemann surfaces

by

ANDREW HAAS

University of Connecticut
Storrs, Connecticut, U.S.A.

One fundamental object of interest in Diophantine approximation is the quantity

$$\nu(x)=\inf\{k\in\mathbf{R}\mid |x-p/q|<k/q^2 \text{ for infinitely many integers } p \text{ and } q\}$$

which describe how well a real number x is approximated by rational numbers. A classical Theorem of A. Markoff states that there is a discrete set of values μ_i decreasing to $\frac{1}{3}$ so that if $\nu(x)>\frac{1}{3}$ then $\nu(x)=\mu_i$ for some i. The theorem also provides a good description of the values μ_i and the numbers x with $\nu(x)>\frac{1}{3}$ [14, 7].

Following the leads of H. Cohn and A. Schmidt we shall investigate geometric structures on hyperbolic Riemann surfaces for which Markoff-like theorems hold. The idea, which is similar to the approach taken by D. Sullivan in [21] is to look at the affinity of a geodesic for the noncompact end of a surface. More precisely, we consider the maximal depth a geodesic travels into a noncompact end. The spectrum of depths has the same structure as Markoff's spectrum with the correspondence given in terms of the geodesics length and topology. The upper discrete part of the spectrum is occupied by the simple closed geodesics and the lower limit value of the discrete spectrum is occupied by the geodesics that are limits of simple closed ones.

The reason for this interaction between the geometry and the number theory becomes apparent when we transfer our attention to a Fuchsian group representing the hyperbolic surface. For example, consider the classical Modular group $\text{Möb}_{\mathbf{Z}}$. The orbit of infinity under the action of $\text{Möb}_{\mathbf{Z}}$ is exactly the set of rational numbers. It is also the set of limit points which are fixed by parabolic transformations in the group. Each parabolic fixed point corresponds, in a sense, to the noncompact end on the quotient surface. As we shall see, the degree to which a number x is approximated by a rational number is directly related to the depth a geodesic with the endpoint x travels

3–868282 *Acta Mathematica* 156. Imprimé le 10 mars 1986

into the noncompact end when projected to the quotient surface. More generally one can look at the approximation of points in the limit set of a Kleinian group by the orbit of a given point. Beardon and Maskit showed in [2] that the geometry of a group is reflected in the approximation properties of points in its limit set.

For a zonal Fuchsian group G and a real number x we define the quantity $\nu_G(x)$ which describes how well the number x is approximated by the orbit of infinity. The geometric results described above can be interpreted to show that for each of the groups we study there is a corresponding Markoff theorem. In other words, there is a discrete set of values $\mu_i(G)$ decreasing to a number $A(G)$ so that if $\nu_G(x)>A(G)$ then $\nu_G(x)=\mu_i(G)$ for some i. This is very similar to Schmidts result in [18]. The numbers x with $\nu_G(x)>A(G)$ are geometrically characterized by the fact that they are precisely the endpoints of lifts of simple closed geodesics from the quotient Riemann surface. As a result we can express the value $\mu_i(G)=\nu_G(x)$ in terms of the length of the associated geodesic. Furthermore, an open leaf of a minimal compactly supported geodesic lamination lifts to a geodesic whose endpoints x and y satisfy $\nu_G(x)=\nu_G(y)=A(G)$.

This gives a new proof of Markoff's theorem as well as a geometric characterization of the Markoff quadratic irrationalities and a subcontinuum of the numbers x with $\nu(x)=\frac{1}{3}$.

The endpoints of a lift of a geodesic may be determined if one has a good description of the geodesics journey around the quotient surface. Such a description is provided by the symbolic dynamics. For our purposes it is therefore desirable to know the symbolic dynamics of simple geodesics. This problem has been studied by Birman and Series in [3] where they present an algorithm for determining simplicity. We will give an explicit description of the symbolic dynamics for simple geodesics on a punctured torus [5]. Following the approach of Birman and Series we then compute the endpoints of lifts of simple geodesics. In the classical case this simplifies to give the continued fraction expansions for the Markoff quadratic irrationalities and the lamination endpoints with $\nu(x)=\frac{1}{3}$.

§ 1. Definitions and statement of results

The recommended reference for the hyperbolic geometry is Beardon's book [1]. We will be working with the upper half plane model $\mathbf{H}$ for hyperbolic space with the Poincaré metric $ds^2=|dz|^2/\mathrm{Im}^2 z$. The length of a curve $\gamma: [a, b]\to\mathbf{H}$ is given by the integral

$$L(\gamma) = \int_a^b \frac{|\dot{\gamma}(t)|}{\mathrm{Im}\,\gamma(t)}\,dt.$$

The orientation preserving isometries of $\mathbf{H}$ are the real Möbius transformations $\text{Möb}_{\mathbf{R}}$ which have the form $z \mapsto (az+b)/(cz+d)$, $a, b, c, d \in \mathbf{R}$ and $ad-bc=1$.

A hyperbolic *geodesic* is a half line or circle in $\mathbf{H}$ orthogonal to the real axis. The endpoints of a geodesic are its *ends*. The angle between two geodesics is measured by the Euclidean angle between their tangent lines at the point of intersection. Two geodesics which share a common endpoint form an angle measuring zero.

A *horocycle* in $\mathbf{H}$ at a point p is an open disc in $\mathbf{H}$ tangent to the real axis if $p \in \mathbf{R}$ or a half plane $\{\operatorname{Im} z > t\}$ if $p = \infty$.

Every hyperbolic Riemann surface N can be represented as the quotient of $\mathbf{H}$ by a Fuchsian group G. Since G is also a group of isometries of $\mathbf{H}$, N inherits a hyperbolic metric by way of the covering projection $\pi: \mathbf{H} \to N$. The length of a curve γ on N, which we denote $L(\gamma)$, is usually calculated by lifting it to a curve in $\mathbf{H}$.

A *geodesic* on N is the image of a geodesic on $\mathbf{H}$ under the covering projection. Let $\tilde{\gamma}$ be a geodesic on $\mathbf{H}$ covering the geodesic γ on N. γ is a *simple* geodesic if $g(\tilde{\gamma}) \cap \tilde{\gamma} = \emptyset$ for all $g \in G$ which do not lie in the subgroup $\operatorname{stab}(\tilde{\gamma})$ whose transformations stabilize $\tilde{\gamma}$. γ is a *closed* geodesic if $\operatorname{stab}(\tilde{\gamma})$ is an infinite cyclic subgroup of G. If $\operatorname{stab}(\tilde{\gamma})$ contains only the identity then γ is an *open* geodesic. We use the term *end of a geodesic* γ to describe the asymptotic behavior of γ on N. More precisely, an end of a geodesic γ_1 is asymptotic to a geodesic γ_2 if they have lifts $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ which have a common endpoint.

We will generally treat a curve on a surface as a point set, thereby minimizing the importance of a particular parametrization. The reader may think of a curve as having a simple $\mathscr{C}^1$ parametrization. When considering the homology or free homotopy class of a closed curve we identify the two possible orientations, thus treating a homology or homotopy class and its inverse as one. This fits with our definition of a geodesic as an unoriented object.

Throughout this paper we will work exclusively with surfaces N which are of finite topological type. This means that there exists a conformal imbedding of N into a compact surface N_0 of the same genus so that the complement of N in N_0 consists of finitely many connected components. These complementary components correspond in a one-to-one fashion with the *noncompact ends* of N. Two types of ends may be distinguished: *finite ends* or *punctures*, which correspond to single point components in $N_0 \setminus N$; and *infinite ends* whose complementary components are not points.

A surface of genus 1 with one noncompact end is called an *h-torus*. A surface of genus 0 with 4 noncompact ends is called a *4x sphere*.

Corresponding to a noncompact end on N is a unique nontrivial free homotopy class of closed curves. A closed curve in this free homotopy class can be contracted to a point when we adjoin to N the component in $N_0 \setminus N$ associated to the given noncompact end. A nontrivial free homotopy class of closed curves contains a unique closed geodesic representative if it does not correspond to a puncture on N. The closed geodesic associated to an infinite noncompact end is simple. We call such a geodesic a *boundary geodesic*.

A sequence of geodesics γ_i on N converges to a geodesic γ if there are lifts $\tilde{\gamma}_i$ of the γ_i and $\tilde{\gamma}$ of γ to $\mathbf{H}$ so that the endpoints of the geodesics $\tilde{\gamma}_i$ converge to the endpoints of $\tilde{\gamma}$. In general a sequence of geodesics may have many limits.

Let $S(N)$, or just S when the meaning is clear, be the set of simple closed geodesics on N which are not boundary geodesics. $\bar{S}$ will denote the set of geodesics which are limits of geodesics in S. It is easily shown that $\bar{S}$ contains only simple geodesics [11, 22].

§1.1. Let G be a Fuchsian group normalized so that the stabilizer of infinity is generated by a transformation of the form $z \mapsto z+t$ for some $0 \neq t \in \mathbf{R}$. Groups with this property are called zonal Fuchsian groups of width t.

For $x \in \mathbf{R}$ define

$$\nu_G(x) = \inf\left\{ k \in \mathbf{R} \,\middle|\, |x - g_n(\infty)| < k/c_n^2 \text{ for infinitely many } \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} = g_n \in G \text{ with } c_i \neq c_j \text{ for } i \neq j \right\}.$$

The set of numbers for which there exist infinitely many group elements g_n with $|x - g_n(\infty)| < k/c_n^2$ for some real k are called *points of approximation* for the group G [1, 2]. Working with the values $\nu_G(x)$ we can delineate a finer structure within the points of approximation which is directly analogous in the cases we study, to the Lagrange spectrum of classical Diophantine approximation. Define the Lagrange spectrum $L(G)$ for the group G to be the set of values $\nu_G(x)$ for all real x. If G is the group $\text{Möb}_{\mathbf{Z}}$ then it is known [17] that $L(G)$ is the classical Lagrange spectrum.

THEOREM 1.1. *Let G be a zonal Fuchsian group of width t representing an h-torus T with a finite noncompact end. Suppose x is not in the G-orbit of infinity. Then $\nu_G(x) > 2/t$ if and only if x is the endpoint of a geodesic $\tilde{\alpha}(x)$ which projects to a simple closed geodesic $\alpha(x)$ on T; the value $\nu_G(x) = (2/t)\coth\frac{1}{2}L(\alpha(x))$. Furthermore, if x is the endpoint of a geodesic $\tilde{\alpha}(x)$ which projects to a geodesic $\alpha(x) \in \bar{S} \setminus S$ then $\nu_G(x) = 2/t$.*

If we let G be the subgroup Γ' of $\text{Möb}_{\mathbf{Z}}$ with generators

$$\begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix} \text{ and } \begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix}$$

then $t=6$ and Theorem 1 implies the classical Markoff theorem. We can also infer that the Markoff quadratic irrationalities ($\nu(x)>\frac{1}{3}$) are exactly the endpoints of lifts of simple closed geodesics from the quotient surface $T_{\mathbf{Z}}=\mathbf{H}/\Gamma'$. Furthermore, a subcontinuum of the numbers x with $\nu(x)=\frac{1}{3}$ are exactly the endpoints of lifts of simple biinfinite geodesics on $T_{\mathbf{Z}}$ which are limits of simple closed geodesics.

Let H be a zonal Fuchsian group representing a $4x$ sphere M with all finite noncompact ends. It is known that there is a finite extension H^* of H which is a Fuchsian group having signature $(0;2,2,2,\infty)$ [1, 18]. Schmidt refers to these groups as extended Fricke groups. The parabolic transformations in H^* are all contained in H. Therefore, if H has width t then so does H^*. Shiengorn was first to observe the geometric connection between H and H^* in the context of the Lagrange spectrum.

THEOREM 1.2. *Let H be a zonal Fuchsian group representing a $4x$ sphere M having all finite noncompact ends. Suppose x is not in the H^* orbit of infinity. Then $\nu_{H^*}(x)>1/t$ if and only if x is the endpoint of a geodesic $\tilde{\alpha}(x)$ which projects to a simple closed geodesic $\alpha(x)$ on M; the value $\nu_{H^*}(x)=(1/t)\coth\frac{1}{4}L(\alpha(x))$. Furthermore, if x is the endpoint of a geodesic $\tilde{\alpha}(x)$ which projects to a geodesic $\alpha(x)\in\bar{S}\setminus S$ then $\nu_{H^*}(x)=1/t$.*

The case where H is the congruence subgroup $\Gamma(3)\subset \text{Möb}_{\mathbf{Z}}$ was studied by Lehner and Shiengorn [12]. Here also the Markoff quadratic irrationalities are exactly the endpoints of lifts of simple closed geodesics from the quotient surface $M_{\mathbf{Z}}=\mathbf{H}/\Gamma(3)$.

The proofs of Theorems 1 and 2 occur in (§ 5).

§ 1.2. The fundamental arguments for this paper are embodied in Theorem 1.3–1.6. In this geometric setting we are able to deal with surfaces that have infinite noncompact ends as well as punctures. Thus Theorems 1.3–1.6 are in a sense generalizations of Theorems 1.1 and 1.2, where instead of working with a covering group we go down to the surface itself.

Given a geodesic α on a hyperbolic surface N with a boundary geodesic γ we define $D(\alpha,\gamma)$, or $D(\alpha)$ when γ is the unique boundary geodesic, to be the infimum of the distance between points on γ and points on α. In other words, $D(\alpha,\gamma)$ is the distance between α and γ.

THEOREM 1.3. *Let T be an h-torus with boundary geodesic γ. A geodesic α on T*

lies on $\bar{S}$ *if and only if* $\coth\frac{1}{4}L(\gamma)\leq\cosh D(\alpha)$. *Equality hold only if* $\alpha\in\bar{S}\setminus S$. *For* $\alpha\in S$ $\cosh D(\alpha)=\coth\frac{1}{2}L(\alpha)\coth\frac{1}{4}L(\gamma)$.

THEOREM 1.4. *Let M be a 4x sphere with four boundary geodesics* $\gamma_1, \gamma_2, \gamma_3, \gamma_4$, *all having equal length. Set* $D(\alpha)=\min\{D(\alpha,\gamma_i)\,|\,i=1,\dots,4\}$. *A geodesic* α *on M lies in* $\bar{S}$ *if and only if* $\coth\frac{1}{2}L(\gamma_i)\leq\cosh D(\alpha)$. *Equality occurs only when* $\alpha\in\bar{S}\setminus S$. *For* $\alpha\in S$,

$$\cosh D(\alpha,\gamma_i)=\coth\frac{L(\gamma_i)}{2}\coth\frac{L(\alpha)}{4}\quad \textit{for } i=1,\dots,4.$$

Let G be a Fuchsian group representing the surface N. There is a one-to-one correspondence between punctures on N and conjugacy classes of maximal parabolic subgroups of G. Let C be a conjugacy class of parabolics corresponding to the puncture p on N. There is a G invariant set of horocycles tangent to the boundary of $\mathbf{H}$ at fixed points of elements in C which projects to a punctured disc on N. This punctured disc is called a horocyclic neighborhood of the puncture p.

Given a geodesic α on N we let $A(\alpha,p)$ denote the area of the largest horocyclic neighborhood of p on N which is disjoint from the geodesic α.

THEOREM 1.5. *Let T be an h-torus with a finite noncompact end. A geodesic* α *on T lies in* $\bar{S}$ *if and only if* $4\leq A(\alpha)$. *Equality occurs only when* $\alpha\in\bar{S}\setminus S$. *For* $\alpha\in S$ $A(\alpha)=4\coth\frac{1}{2}L(\alpha)$.

THEOREM 1.6. *Let M be a 4x sphere with only finite noncompact ends* p_1, p_2, p_3, p_4. *Set* $A(\alpha)=\min\{A(\alpha,p_i)\,|\,i=1,\dots,4\}$. *A geodesic* α *on M lies in* $\bar{S}$ *if and only if* $2\leq A(\alpha)$. *Equality occurs only when* $\alpha\in\bar{S}\setminus S$. *For* $\alpha\in S$ $A(\alpha,p_i)=2\coth\frac{1}{4}L(\alpha)$ *for* $i=1,\dots,4$.

The proofs of Theorems 1.3–1.6 all follow the same general outline. To begin one proves the identity for simple closed geodesics (§ 2). This is done by dissecting the surface along an appropriately chosen set of curves. The dissection results in a plane hyperbolic polygon which is then analysed using plane hyperbolic geometry. This is extended to all of $\bar{S}$ by taking limits (§ 4.1).

Geodesics which are not simple, and hence not in $\bar{S}$, are shown to contain arcs which bound either a geodesic monogon or bigon about a noncompact end. Again, after dissecting the surface along these curves we use plane hyperbolic geometry to show that the existence of such a configuration forces the geodesic to make sufficiently deep incursions into a noncompact end. (§ 3.1).

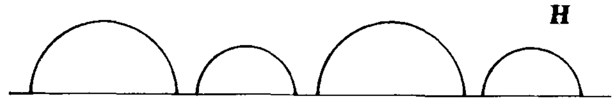

Fig. 1.1 (a)

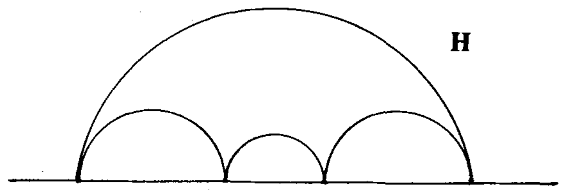

Fig. 1.1 (b)

It then remains to consider the simple geodesics which do not belong to $\tilde{S}$. In (§ 4) we determine exactly which geodesics fall into this category. It is then easy to get a rough estimate of the depth of their incursions into a noncompact end.

The different pieces of the argument will be tied together in § 4.1.

§ 1.3. Let G be a Fuchsian group representing an h-torus T. There is a fundamental domain F for G with four disjoint geodesics along its boundary, as illustrated in Figures 1.1 (a) and (b). If T has a finite noncompact end then F is an ideal quadrilateral in $\mathbf{H}$. Transformations A and B pair opposite sides of F and generate G. Let $\mathcal{S}=\{a, a^{-1}, b, b^{-1}\}$ be the set of sides of F where $A(a)=a^{-1}$ and $B(b)=b^{-1}$.

A function $\sigma: \mathbf{Z}\to\mathcal{S}$ is *freely reduced* if $\sigma(n)$ is different from $\sigma(n+1)^{-1}$ and $\sigma(n-1)^{-1}$ for all $n\in\mathbf{Z}$. Define an equivalence relation on the set of functions $\sigma: \mathbf{Z}\to\mathcal{S}$ where σ is equivalent to σ' if either $\sigma(n)=\sigma'(n+k)$ or $\sigma(n)=\sigma'(-n+k)$ for all $n\in\mathbf{Z}$ and some $k\in\mathbf{Z}$.

A geodesic γ on T which does not terminate at the noncompact end determines a unique equivalence class of freely reduced functions $\sigma: \mathbf{Z}\to\mathcal{S}$ (§ 6). We call this equivalence class the *symbolic dynamics* of γ and denote it by $\mathrm{Dyn}(\gamma)$. This approach is essentially the same as the one in [3] where γ is represented by a biinfinite word in the symbols of $\mathcal{S}$. It is shown there that each equivalence class of freely reduced functions determines a geodesic γ on T.

Let $\sigma: \mathbf{Z}\to\mathcal{S}$ be freely reduced. The restriction of σ to the interval $[j, k]$ determines a *maximal block* if $\sigma(n)=s$ for $j\leq n\leq k$ and $\sigma(j-1), \sigma(k+1)\neq s$. There is a natural ordering of the maximal blocks in σ. We may then write σ in condensed form as a function $\Sigma: \mathbf{Z}\to W$, where $W=\{a^m, b^m \mid m\in\mathbf{Z}\}$. $\Sigma(n)$ is defined to be s^m where $\sigma=s$ on the nth maximal block determined by the interval $[j, k]$ with $m=k-j+1$. We will generally use condensed functions Σ to represent symbolic dynamics.

For $x \in \hat{\mathbf{R}}$ and $c \in \mathbf{R}$ define the function $q(x, c)\colon \mathbf{Z} \to \mathbf{Z}$ by

$$q(x, c)(n) = [(n+1)x + c] - [nx + c]$$

where [ ] denotes the greatest integer function.

When $x \in \hat{\mathbf{R}}$ and $c \neq mx+k$ for any $m, k \in \mathbf{Z}$, let $\Sigma(x, c)\colon \mathbf{Z} \to W$ be defined by

$$\begin{aligned}
\Sigma(x, c)(n) &= \begin{cases} a^{\operatorname{sgn}(x)}, & n = 2m,\ m \in \mathbf{Z} \\ b^{\operatorname{sgn}(x)\, q(x, c)(m)}, & n = 2m+1,\ m \in \mathbf{Z} \end{cases} \quad \text{when } 1 \leqslant |x| < \infty \\
&= \begin{cases} b^{\operatorname{sgn}(x)}, & n = 2m,\ m \in \mathbf{Z} \\ a^{\operatorname{sgn}(x)\, q(1/x, -c/x)(m)}, & n = 2m+1,\ m \in \mathbf{Z} \end{cases} \quad \text{when } 0 < |x| < 1 \\
&= \begin{cases} b & \text{when } x = \infty \\ a & \text{when } x = 0. \end{cases}
\end{aligned}$$

When x is irrational and $c = mx+k$ for $m, k \in \mathbf{Z}$

$$\Sigma^1(x, c)(n) = \begin{cases} \Sigma(x, c)(n), & n \neq 1 & \\ b^{\operatorname{sgn}(x)([x]-1)}, & n = -1 & 1 \leqslant x < \infty \\ a^{\operatorname{sgn}(x)([1/x]-1)}, & n = -1 & 0 \leqslant x < 1 \end{cases}$$

and

$$\Sigma^2(x, c)(n) = \begin{cases} \Sigma(x, c)(n), & n \neq 1 & \\ b^{\operatorname{sgn}(x)([x]-1)}, & n = -1 & 1 \leqslant x < \infty \\ a^{\operatorname{sgn}(x)([1/x]-1)}, & n = -1 & 0 \leqslant x < 1 \end{cases}$$

A *geodesic lamination* on a hyperbolic surface N is a closed set of disjoint simple geodesics on N. We denote by $\mathscr{GL}$ the set of all laminations on N and by $\mathscr{GL}_0$ the subset of compactly supported laminations. A lamination is *minimal* if it contains no proper sublaminations.

THEOREM 1.7. *Let γ be a simple geodesic in $\tilde{S}(T)$ which belongs to some minimal geodesic lamination on T. Then* $\operatorname{Dyn}(\gamma)$ *is represented by one of the functions $\Sigma(x, c)$, $\Sigma^1(x, c)$ or $\Sigma^2(x, c)$ for some $(x, c) \in \mathbf{R}^2$.*

It will become clear in § 6 that the seemingly restrictive hypothesis of Theorem 1.7 actually allows for a description of the symbolic dynamics of all simple geodesics on T.

Let $\xi = [a_0, a_1, \ldots]$ be the number with continued fraction expansion

$$a_0+\cfrac{1}{a_1+\cfrac{1}{a_2+\dots}}.$$

Using Dickenson notation in [7] we may write the continued fraction expansion of a number ξ in the form $(n_1)_{m_1}(n_2)_{m_2}(n_3)_{m_3}\dots$, where each $(n_i)_{m_i}$ denotes a repeating block of m_i consecutive occurances of the integer n_i in the continued fraction expansion of ξ.

THEOREM 1.8. *ξ is the endpoint of a lift of a simple geodesic on $T_{\mathbf{Z}}$ which does not terminate at the puncture if and only if $\xi=[b_0, b_1, \dots, b_k, a_n, a_{n+1}, \dots]$ or $[a_n, a_{n+1}, \dots]$ where $[a_0, a_1, a_2, \dots]$ is one of the following:*

(1) 1_∞,

(2) $1_{2q(0)-1}2_2 1_{2q(1)-2}2_2\dots 2_2 1_{2q(n)-2}2_2\dots$ *where $q(n)=q(x, c)(n)$ with $1\leq x<\infty$ and $c\neq mx+k$ for any $m, k\in\mathbf{Z}$.*

COROLLARY (Markoff). *$\nu(\xi)>\frac{1}{3}$ if and only if $\xi=[b_0, b_1, \dots, b_2, a_n, a_{n+1}, \dots]$ or $[a_n, a_{n+1}, \dots]$ where $[a_0, a_1, \dots]$ is either 1_∞ or $1_{2q(0)-1}2_2 1_{2q(1)-2}\dots 2_2 1_{2q(n)-2}2_2\dots$ with $q(n)=q(x, c)(n)$ for some $x\in\mathbf{Q}$, $1\leq x<\infty$, and c irrational.*

§ 2. Simple closed geodesics

In this section we will derive the formulas for $D(\alpha)$ and $A(\alpha)$ when α is a simple closed geodesic. There are somewhat different derivations of these formulas due to J. P. Matelski. The following two theorems from plane hyperbolic geometry play a crucial role. Their proofs may be found in [1].

THEOREM 2.1. *Refer to the hyperbolic hexagon illustrated in Figure 2.1 where side lengths are labeled.*

$$\cosh b_1 \sinh a_2 \sinh a_3 = \cosh a_1 + \cosh a_2 \cosh a_3.$$

THEOREM 2.2. *Refer to Figure 2.2.*

$$\cosh a \cosh c + \cos\varphi = \sinh a \cosh b \sinh c.$$

When dealing with hyperbolic polygons we adopt the convention that a side that has its length labeled by a Roman lower case letter will lie on a geodesic labeled by the corresponding Greek letter. Also, the side of a polygon determined by a geodesic α will be denoted by $s(\alpha)$.

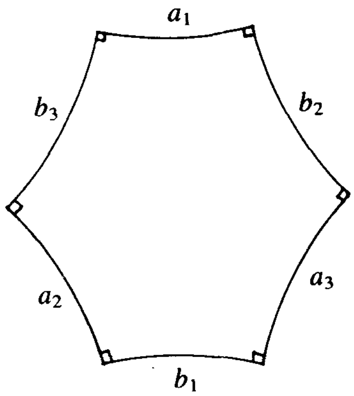

Fig. 2.1

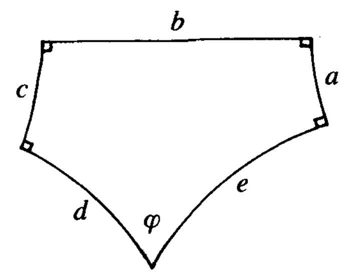

Fig. 2.2

LEMMA 2.1. *Let θ be the hyperbolic octagon illustrated in Figure* 2.3 (a). *Suppose that the side lengths and angles of θ satisfy $d_1=d_2$, $a=c_1+c_2$, $b_1=b_2$, and $\nu_1+\mu_2=\nu_2+\mu_1=\pi$. Then*

$$\cosh d_1 = \coth\frac{a}{2}\coth\frac{e}{4}.$$

Also, the minimal length arc joining α and ε has length d_1.

Proof. There is a unique orientation preserving isometry h which fixes the geodesic α, has translation length a, and takes $s(\beta_1)$ to $s(\beta_2)$. Since μ_2 and ν_1 are complementary angles h also maps γ_1 onto γ_2.

We need to see that the three geodesics γ_1, γ_2, and α never intersect one another either in $\mathbf{H}$ or on its boundary. Suppose $\gamma_1 \cap \alpha \neq \emptyset$. Then since h fixes α and maps γ_1 to γ_2, $\gamma_2 \cap \alpha \neq \emptyset$. The three geodesics α_1, β_1 and γ_1 then bound a triangle in which two interior angles are complementary to μ_1 and μ_2; consequently, $\mu_1+\mu_2>\pi$. By similar reasoning $\nu_1+\nu_2>\pi$. This implies that $\nu_1+\nu_2+\mu_1+\mu_2>2\pi$ contrary to the hypothesis. It then follows easily that $\gamma_1 \cap \gamma_2=\emptyset$.

We are now able to define the common perpendicular β_1^* of γ_1 and α, and the common perpendicular β_2^* of γ_2 and α. By replacing the side defining geodesics β_1 and β_2 of θ by the geodesics β_1^* and β_2^* we get a right octagon θ^* (see Figure 2.3 (b)).

Since the common perpendicular of two geodesics is unique $h(\beta_1^*)=\beta_2^*$ and $b_1^*=b_2^*$. It also follows that the translation length of h is equal to a^*, and thus $a=a^*$.

Since their endpoints agree on γ_2 the segments $s(\gamma_2)\cup h(s(\gamma_1))$ and $s(\gamma_2^*)\cup h(s(\gamma_1^*))$ are identical; which shows that $c_1+c_2=c_1^*+c_2^*$.

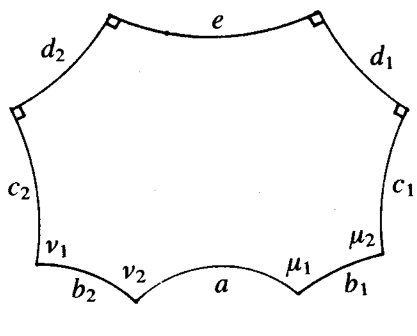

Fig. 2.3 (a)

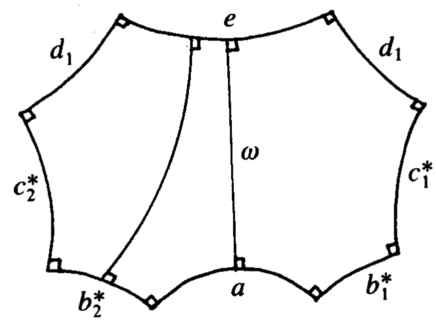

Fig. 2.3 (b)

Let ω be the common perpendicular of α and ε. Because θ is a right polygon ω joins the sides $s(\alpha)$ and $s(\varepsilon)$, and divides θ into right hexagons H_1 and H_2. Three side lengths determine a right hexagon. Observing the order of correspondence between the sides of equal length $L(\omega), d_2, b_2^*$ and $L(\omega), d_1, b_1^*$ we see that H_1 is congruent to H_2 by the reflection in ω. It follows that $c_1^*=c_2^*=a/2$, and ω bisects both $s(\alpha)$ and $s(\varepsilon)$. Applying Theorem 2.1 to H_1 results in the identity

$$\cosh d_1 \sinh\frac{e}{2}\sinh c_1^* = \cosh\frac{a}{2} + \cosh\frac{e}{2}\cosh c_1^*.$$

Substituting $a/2$ for c_1^* and then solving for $\cosh d_1$ gives

$$\cosh d_1 = \coth\frac{a}{2}\left(\cosh\frac{e}{2} + \coth\frac{e}{2}\right) = \coth\frac{a}{2}\coth\frac{e}{4}$$

as claimed.

To complete the argument we note that H_1 is invariant under reflection in the common perpendicular of ε and β_1^*. Q.E.D.

THEOREM 2.3. *Let T be an h-torus with boundary geodesic γ. For $\alpha \in S(T)$, $\cosh D(\alpha) = \coth \frac{1}{2}L(\alpha)\coth\frac{1}{4}L(\alpha)$.*

Proof. Choose a minimal length geodesic arc δ joining α to γ. δ meets both α and γ at right angles. Pick a point p on α and a closed noncontractible curve c beginning and ending at p which meets α only at p and is disjoint from δ and γ. Let β be the closed geodesic segment in the homotopy class of the curve c with the same fixed endpoint p. β is a closed curve on T although it will not in general be a closed geodesic according to our definition (see Figure 2.4). Nonetheless β has the important property of being

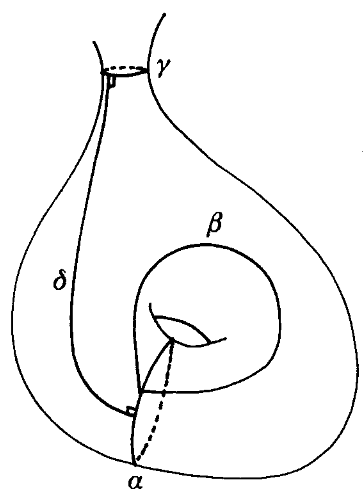

Fig. 2.4

disjoint from both γ and δ and meeting α in a single point. This is due to the fact that if for example β intersected δ it would be in two points and together they would bound a simply connected region on T. This sort of intersection (non-geometric) cannot occur between geodesics.

The region on T bounded by the geodesics α, γ, β and δ is an octagon which we may realize, by lifting, as an octagon θ in hyperbolic space. θ is easily seen to satisfy the hypothesis of Lemma 2.1; hence,

$$\cosh L(\delta) = \coth \frac{L(\alpha)}{2} \coth \frac{L(\gamma)}{4}.$$

As $L(\delta)=D(\alpha)$ the theorem is proven. Q.E.D.

It is interesting to observe at this point that there is another geodesic arc δ' on T which realizes the minimal distance $D(\alpha)$ between α and γ. This follows from Lemma 2.1 by letting δ' be the projection to T of the common perpendicular joining the lifts of α and γ bounding θ. It is also clear that δ and δ' approach α from opposite sides.

Let P be a polygon with a vertex v along the boundary of $\mathbf{H}$. We will be considering the region of P lying inside a horocycle tangent to the boundary of $\mathbf{H}$ at the vertex v. If s is a side of P define $A_v(s)$ to be the area of the largest horocyclic neighborhood of v in P disjoint from s.

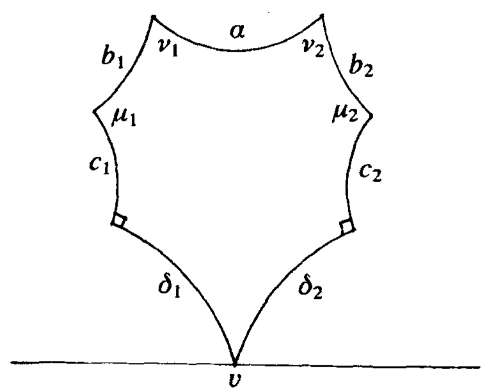

Fig. 2.5 (a)

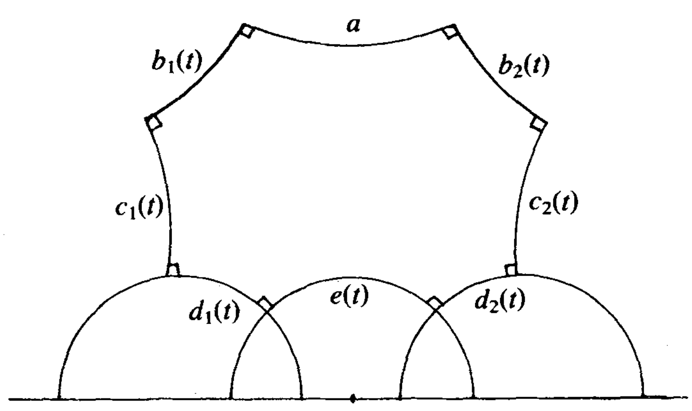

Fig. 2.5 (b)

LEMMA 2.2. *Let P be the septagon illustrated in Figure 2.5 (a) with the vertex v on the real axis. Suppose that $c_1+c_2=a$, $b_1=b_2$, $\mu_1+\mu_2=\nu_1+\nu_2=\pi$ and $A(s(\gamma_1))=A(s(\gamma_2))$. Then*

$$A(s(\gamma_1)) = A(s(\gamma_2)) = 4\coth\frac{a}{2} = A(s(\alpha)).$$

Proof. There is no loss of generality in supposing that the geodesic ω eminating from v and perpendicular to α bisects $s(\alpha)$. This can be achieved by varying the polygon with a parabolic isometry of $\mathbf{H}$ which fixes v. Using the methods of Lemma 2.1 we may further suppose that $\mu_1=\mu_2=\nu_1=\nu_2=\pi/2$. Then since reflection in ω maps δ_1 to δ_2, $s(b_1)$ to $s(b_2)$, and leaves $s(a)$ invariant it follows that $c_1=c_2$.

We construct open polygons $P(t)$ depending on a real parameter t so that $b_i(t)=b_i+1/t$ (see Figure 2.5 (b)). To do this first fix the geodesics β_1, β_2 and α. Let $\gamma_i(t)$ be the geodesic orthogonal to β_i with $b_i(t)=b_i+1/t$. Then define $\delta_i(t)$ to be a geodesic orthogonal to $\gamma_i(t)$ so that $c_i(t)=c_i$. $\delta_1(t)$ and $\delta_2(t)$ are clearly disjoint. Adjoining the common perpendicular $\varepsilon(t)$ of $\delta_1(t)$ and $\delta_2(t)$ to $P(t)$ gives a right octagon $\theta(t)$. Since $P(t)$ and hence $\theta(t)$ are invariant under reflection in ω, $d_1(t)=d_2(t)$ and $\theta(t)$ satisfies the hypothesis of Lemma 2.1.

It follows that

$$\cosh d_1(t) = \coth\frac{a}{2}\coth\frac{e(t)}{4}$$

and the distance from $s(e(t))$ to α is $d_1(t)$.

From the definitions it is seen that $\theta(t)$ converges uniformly to P on compact sets. The neighborhoods $N(t)$ of the geodesic $s(\varepsilon)$ in $\theta(t)$ of width $d_1(t)$ converges to the horocyclic neighborhood of v in P with the desired area $A(s(\gamma_1))$. This shows that $A(s(\alpha))=A(s(\gamma_1))$.

The area $A(N(t))$ converges to $A(s(\gamma_1))$, which leaves us the task of computing $\lim_{t\to\infty} A(N(t))$.

$$\begin{aligned} A(N(t)) &= e(t)\sinh d_1(t) \\ &= e(t)(\cosh^2 d_1(t)-1)^{1/2} \\ &= e(t)\coth\frac{a}{2}\left[\coth^2\frac{e(t)}{4}-\tanh^2\frac{a}{2}\right]^{1/2}. \end{aligned}$$

Since $\lim_{t\to\infty} e(t)=0$, taking limits gives $A(s(\gamma_1))=\lim_{t\to\infty} A(N(t))=4\coth a/2$. Q.E.D.

THEOREM 2.4. *Let T be an h-torus with a finite noncompact end. For $\alpha \in S(T)$, $A(\alpha)=4\coth\frac{1}{2}L(\alpha)$.*

Proof. Choose a simple geodesic ray δ which meets α at one end in a right angle and heads out towards the puncture in the other direction. As in the proof of Theorem 2.3 let β be a geodesic arc beginning and ending at a point p on α, meeting α only at p, and disjoint from δ (see Figure 2.6).

The geodesics α, β, and δ bound a simply connected region on T which lifts to a hyperbolic 7-gon P in H with one vertex v on the boundary.

A horocyclic neighborhood on T disjoint from α projects to a horocyclic neighborhood of v in P. Applying Lemma 2.2 gives the desired result. Q.E.D.

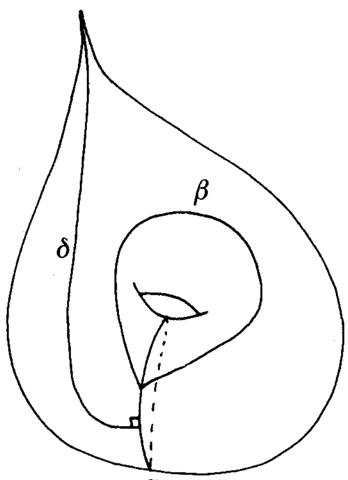

Fig. 2.6

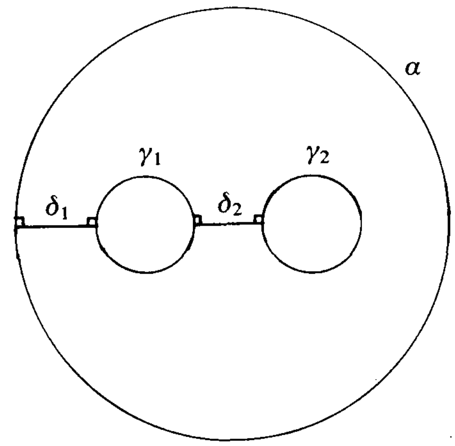

Fig. 2.7

Again we may observe from the lemma that the boundary of the horocyclic neighborhood of area $4\coth\frac{1}{2}L(\alpha)$ touches α exactly twice from opposite sides.

THEOREM 2.5. *Let M be a 4x sphere with four equal length boundary geodesics $\gamma_1, \gamma_2, \gamma_3, \gamma_4$. The distance from $\alpha \in S(M)$ to a boundary geodesic γ_i is*

$$d(\alpha, \gamma_i) = \coth\frac{L(\gamma_i)}{2}\coth\frac{L(\alpha)}{4}.$$

Proof. Since α is a closed curve on the planar surface M α must be a dividing curve. α is not a boundary geodesic so both complementary components are two holed discs D_1 and D_2.

Suppose that γ_1 and γ_2 lie on D_1. Let δ_1 be the minimal length perpendicular joining α to γ_1. Let δ_2 be a geodesic arc in the complement of δ_1 orthogonal to both γ_1 and γ_2 (see Figure 2.7).

The region of D_1 in the complement of δ_1 and δ_2 is simply connected and lifts to a right hexagon H in $\mathbf{H}$. H is easily seen to satisfy the hypothesis of Lemma 2.1 and the theorem follows.

LEMMA 2.3. *Let P be the pentagon illustrated in Figure* 2.8 (a) *with three vertices v_1, v_2, and v_3 on the real axis. Suppose that P is mapped onto itself by reflection in the*

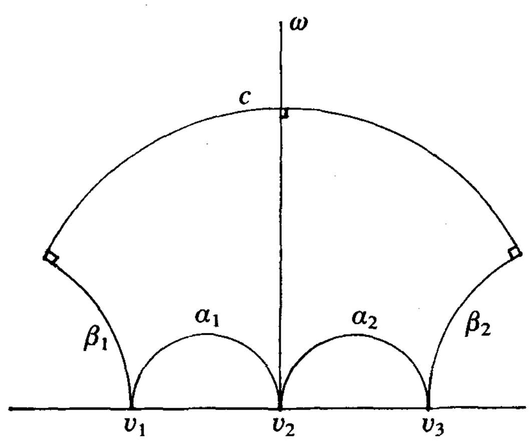

Fig. 2.8 (a)

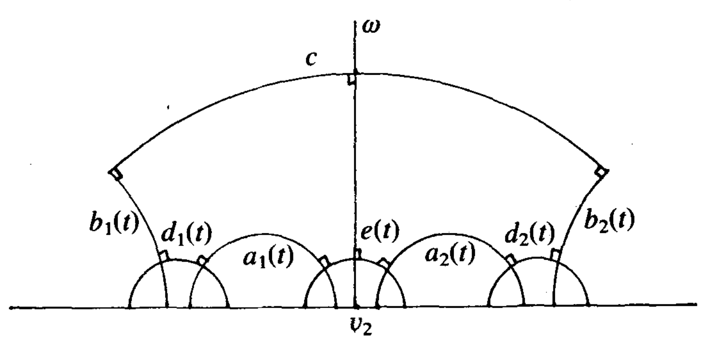

Fig. 2.8 (b)

geodesic ω eminating from v_2 perpendicular to γ. Then the largest horoball neighborhood of v_2 in P not meeting γ has area $A_{v_2}(\gamma)=2\coth\frac{1}{4}c$.

Proof. Consider right octagons $\theta(t)$ as illustrated in Figure 2.8 (b). They are constructed by replacing the geodesics α_i with geodesics $\alpha_i(t)$ lying on circles of smaller radius so that the adjoined perpendicular segments satisfy $d_1(t)+d_2(t)=e(t)=1/t$. We further stipulate that this be done so that $\theta(t)$ is preserved by reflection in ω.

It follows from the final statement of Lemma 2.1 that $b_i(t)=\omega(t)$. As in Lemma 2.2 the neighborhood $N(t)$ of $s(\varepsilon(t))$ in $\theta(t)$ of width $a_1(t)$ converges to the desired cusp neighborhood of v. As this convergence is uniform on compact sets we see that $A(N(t))$ converges to $A_v(\gamma)$.

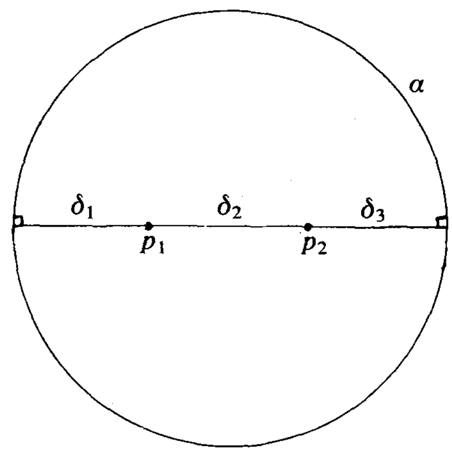

Fig. 2.9

$$\begin{aligned} A(N(t)) &= e(t)\sinh b_1(t) \\ &= e(t)\left(\cosh^2 b_1(t)-1\right)^{1/2} \\ &= e(t)\left(\coth^2\frac{e(t)}{2}\coth^2\frac{c}{4}-1\right)^{1/2}. \end{aligned}$$

Hence $\lim_{t\to\infty} A(N(t))=2\coth\frac{1}{4}c$ as required. Q.E.D.

THEOREM 2.6. *Let M be a $4x$ sphere with all puncture type boundary components. For $\alpha\in S(M)$,*

$$A(\alpha)=2\coth\frac{L(\alpha)}{4}=A_{\varrho_i}(\alpha)$$

for all punctures ϱ_i on M.

Proof. The geodesic α divides M into twice punctured discs D_1 and D_2. Let ϱ_1 and ϱ_2 denote the punctures on D_1. We learn in complex analysis that D_1 can be mapped conformally into the unit disc Δ in $\mathbf{C}$ so that ϱ_1 and ϱ_2 both lie on the real axis with $\varrho_1=-\varrho_2$. The reflection $z\mapsto\bar z$ leaves the image of D_1 in Δ invariant. Hence there is an orientation reversing isometry of D_1 onto itself which fixes three geodesics: δ_1 joining ϱ_1 to α, δ_2 joining ϱ_1 and ϱ_2, and δ_3 joining ϱ_2 to α (see Figure 2.9).

The complement of δ_1 and δ_2 on D_1 is simply connected and lifts to a hyperbolic pentagon P with three vertices along the boundary. δ_3 lifts to a geodesic $\tilde\delta_3$ in P joining α to the ideal vertex corresponding to the puncture ϱ_2. The orientation reversing isometry of D_1 is covered by the reflection in δ_3. It follows that reflection in δ_3 leave invariant the pentagon P.

The hypotheses of Lemma 2.3 are satisfied and the theorem follows. Since the result only depends on $L(\alpha)$ we see that $A(\alpha)=A_{\varrho_i}(\alpha)$ for each $i=1,2,3,4$. Q.E.D.

§ 3. Self intersecting geodesics

Let P be a region on the hyperbolic surface N lying in the complement of some number of geodesic arcs and geodesic rays. If P contains a noncompact end of N and has infinite cyclic fundamental group then P is a geodesic polygon bounding a noncompact end of N. A vertex of P is a point on the boundary of P at which two of the geodesic arcs bounding P intersect. An ideal vertex of P corresponds to two rays on the boundary of P with asymptotic ends.

A geodesic γ on N is said to bound a *monogon* on N if the complement of some piece of γ is a geodesic polygon with one vertex or ideal vertex bounding a noncompact end of N. γ is said to bound a *bigon* on N if there exists a geodesic polygon P with either two vertices or one vertex and one ideal vertex bounding a noncompact end on a hyperbolic surface R and an isometric immersion of P into N mapping the boundary of P into γ. In all cases under consideration the polygon P will lie on a covering of the surface N and the immersion will be induced by the covering projection. The determination of this covering is made explicit in the arguments and at a later stage it will justify our treating the bigon as an embedded geodesic polygon.

Let $\tilde{T}=\mathbf{C}\setminus\mathbf{Z}[i]$, the plane with integer lattice points deleted. The group Γ containing transformations of the form $z\mapsto z+m+in$ for $m,n\in\mathbf{Z}$, acts discontinuously on $\tilde{T}$. The quotient $\tilde{T}/\Gamma$ is an h-torus T_0. A straight line $l(r)$ in $\tilde{T}$ with slope $r\in\hat{\mathbf{Q}}=\mathbf{Q}\cup\{\infty\}$ is invariant under transformations $z\mapsto z+m+in$ with $n/m=r$. All other transformations in Γ map $l(r)$ disjointly. $l(r)$ therefore projects to a simple closed curve on T_0 which is freely homotopic to a simple closed geodesic $\lambda(r)$. Moreover, $\lambda(r)$ depends only on the value r (see § 6 or [9]).

Consider the collection $\mathscr{L}_h(\mathscr{L}_v)$ of horizontal (vertical) lines in $\mathbf{C}$ which pass through points in $\mathbf{Z}[i]$. The reflection in a line l in either $\mathscr{L}_h$ or $\mathscr{L}_v$ is an anticonformal homeomorphism of $\tilde{T}$ and consequently a hyperbolic isometry. Since the arcs of l that lie in $\tilde{T}$ are fixed pointwise by the reflection they are geodesics in the hyperbolic metric on $\tilde{T}$. By the same reasoning horizontal lines with half integer imaginary part and vertical lines with half integer real part are geodesics in $\tilde{T}$.

The collection of lines $\mathscr{L}_\#=\mathscr{L}_h\cup\mathscr{L}_v$ divides $\mathbf{C}$ up into squares with vertices in $\mathbf{Z}[i]$.

Given a geodesic $\tilde{\gamma}$ in $\tilde{T}$ which does not terminate at a puncture point (in $\mathbf{Z}[i]$) and points p and q on $\tilde{\gamma}$ there is a clearly defined sequence of squares in $\tilde{T}$ describing the path from p to q. This gives a loose characterization of the symbolic dynamics of the arc $\tilde{\gamma}(p, q)$. We shall delve more deeply into this matter in §6.

Let $\tilde{\sigma}(r, s)$ be another arc of a geodesic $\tilde{\sigma}$ in $\tilde{T}$ with endpoints r and s. Suppose that the path $\tilde{\gamma}(p, q)$ traverses exactly the same sequence of n squares in $\tilde{T}$ as the path $\tilde{\sigma}(r, s)$. An equivalent formulation of Lemma 1.1 in [4] is

LEMMA 3.1. *There exist constants k, c, and a so that a subarc of $\tilde{\sigma}(r, s)$ of hyperbolic length kn is contained within a tubular neighborhood of $\tilde{\gamma}(p, q)$ of width ce^{-an}.*

PROPOSITION 3.1. *Let γ be a closed self intersecting geodesic on an h-torus T. Then γ must bound a monogon or a bigon on T.*

Proof. There is a homeomorphism $g: T \to T_0$ mapping the given h-torus onto T_0. Clearly, γ bounds a bigon or a monogon on T if and only if the geodesic on T_0 freely homotopic to $g(\gamma)$ also bounds a bigon or a monogon. It is therefore no loss of generality to restrict our attention to the particular h-torus T_0.

We assume γ on T_0 does not bound a monogon and show that γ must then bound a bigon. It will suffice to demonstrate the existence of two lifts of γ to $\tilde{T}$ which together bound a bigon on $\tilde{T}$.

First consider the case where γ lifts to a geodesic $\tilde{\gamma}$ in $\tilde{T}$ which is not a closed geodesic. Then there is a nontrivial maximal cyclic subgroup $\Gamma_0 \subset \Gamma$ generated by $z \mapsto z + m_0 + in_0$ for some integers m_0 and n_0 which stabilizes every lift of γ to $\tilde{T}$.

It follows from [8] that there is a homeomorphism $f: T_0 \to T_0$ taking the geodesic $\lambda(n_0/m_0)$ onto a closed curve which is freely homotopic to $\lambda(0)$. The geodesic on T_0 in the free homotopy class of the closed curve $f(\gamma)$ will bound a bigon exactly when γ does. Hence, there is no loss of generality if we suppose that $n_0 = 0$.

Let α be a geodesic on T_0 which does not terminate at the puncture, and which is disjoint from the projection of the lines in $\mathcal{L}_h$ to T_0. Lift α to a geodesic $\tilde{\alpha}$ in $\tilde{T}$ lying in the region Ω between the lines $l_0 = \{\operatorname{Im} z = 0\}$ and $l_1 = \{\operatorname{Im} z = 1\}$. Since $\tilde{\alpha}$ realizes only geometric intersections with the geodesic arcs of $\mathcal{L}_v$ in T, α may only cross an arc in $\mathcal{L}_v$ once. It follows then that $\tilde{\alpha}$ cannot make a "u-turn" in Ω, and consequently it must cross each arc of $\mathcal{L}_v$ in Ω exactly once. Thus the sequence of squares traversed by any subarc of α agrees with the sequence traversed by a corresponding subarc of the geodesic $l_{1/2} = \{\operatorname{Im} z = \frac{1}{2}\}$. By Lemma 3.1 it must be that $\tilde{\alpha} = l_{1/2}$ and thus $\alpha = \lambda(0)$.

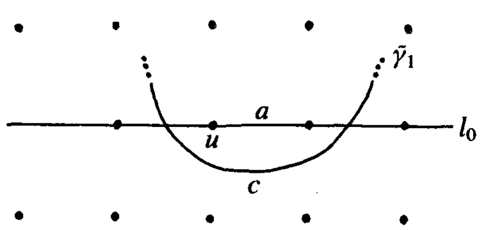

Fig. 3.1

Since γ is distinct from $\lambda(0)$ it has a lift $\tilde{\gamma}$ which intersects l_0. Both l_0 and $\tilde{\gamma}$ remain invariant under $z \mapsto z+m_0$ so l_0 and $\tilde{\gamma}$ must actually intersect in infinitely many points. Nevertheless, there are only finitely many intersections occuring between γ and the projection of l_0 on T_0. $\tilde{\gamma}$ divides $\tilde{T}$ into two unbounded regions, one above and the other below $\tilde{\gamma}$. Let $u+iv$ be a lattice point which does not lie in the unbounded region below $\tilde{\gamma}$, and further suppose that its imaginary part is minimal among all such points. A minimum must exist since γ is compact on T_0. Set $h(z)=z-iv$. Then there is an arc a of l_0 containing the integer u whose endpoints are also the endpoints of a simple arc c of the geodesic $h(\tilde{\gamma})=\tilde{\gamma}_1$. Together a and c bound a region below l_0 which is disjoint from $\mathbf{Z}[i]$ (see Figure 3.1).

Let a_1 be an arc of l_0 that satisfies the following:

(1) There is a lift $\tilde{\gamma}_1$ of γ containing an arc c_1 lying below l_0 and meeting a_1 at its endpoints.

(2) a_1 contains at least one lattice point.

(3) The region bounded by a_1 and c_1 does not contain any lattice points.

(4) a_1 contains a minimal number of lattice points among all arcs satisfying the first three properties.

Similarly we may choose an arc a_2 and a geodesic $\tilde{\gamma}_2$ containing an arc c_2 so that properties (1)–(4) hold with the modifications that the region bounded by a_2 and c_2 lie above l_0.

Let n_1 be the largest integer on a_1 and let n_2 be the smallest integer on a_2. Set $\tau(z)=z+n_2-n_1$. The arc $a=a_2 \cap \tau(a_1)$ contains exactly the integer n_2. We will show that $\tilde{\gamma}_2$ and $\tau(\tilde{\gamma}_1)$ together bound a bigon containing the lattice point n_2.

We first suppose that one of the endpoints of c_2 on l_0 lies between the two endpoints of $\tau(c_1)$ and that one of the endpoints of $\tau(c_1)$ lies between the endpoints of c_2. This is the situation illustrated in Figure 3.2 (a). Let us follow $\tilde{\gamma}_2$ beyond c_2 into the region below l_0 bounded by $\tau(c_1)$. The important thing to notice is that $\tilde{\gamma}_2$ must exit

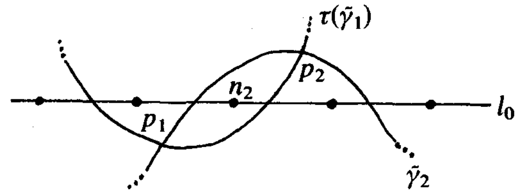

Fig. 3.2 (a)

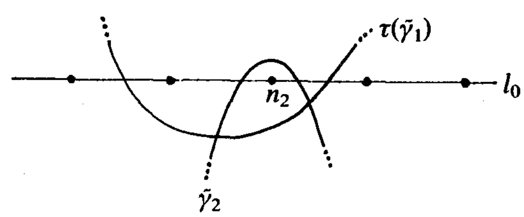

Fig. 3.2 (b)

from this region by crossing $\tau(c_1)$ at a point p_1. If $\tilde{\gamma}_2$ were rather to cross $\tau(a_1)$ first then the subarc of $\tau(a_1)$ lying between this crossing and the right endpoint of a_2 would satisfy the first three defining properties of a_1, but this arc then contains fewer lattice points than a_1 which contradicts the definition. For the same reason the extension of $\tau(\gamma_1)$ beyond $\tau(c_1)$ into the region bounded by a_2 and c_2 must exit by crossing c_2 at a point p_2. The segments of $\tilde{\gamma}_2$ and $\tau(\tilde{\gamma}_1)$ joining p_1 to p_2 form a bigon containing exactly the point n_2.

Without loss of generality the remaining possibility is that both endpoints of c_2 lie between the endpoints of $\tau(c_1)$ (Figure 3.2 (b)). Following $\tilde{\gamma}_2$ beyond c_2 from both sides we see, as above, that $\tilde{\gamma}_2$ must exit from the region bounded by $\tau(a_1)$ and $\tau(c_1)$ by crossing $\tau(c_1)$. The bigon is evident.

The remaining case, where $\tilde{\gamma}$ is a closed geodesic, is now elementary. We may suppose $\tilde{\gamma}$ meets l_0 and as above, find arcs of l_0 satisfying properties (1)–(4). The argument then proceeds without any alterations. Q.E.D.

By definition a geodesic γ *spirals around a closed geodesic* if one of the endpoints of a lift $\tilde{\gamma}$ of γ is a fixed point of a hyperbolic transformation in the covering group. The fixed axis of that hyperbolic projects to the closed geodesic that γ spirals around.

THEOREM 3.1. *Let γ be a compactly supported self intersecting geodesic or geodesic ray on an h-torus, and suppose that γ does not spiral around the boundary geodesic if one exists. Then either γ bounds a monogon, γ bounds a bigon, or there is a simple closed geodesic α which along with γ bounds a bigon.*

Proof. The case where γ is a closed geodesic has already been treated in Proposition 3.1. It will suffice to argue the case where γ is a geodesic ray. Again we suppose that γ does not bound a monogon and show that one of the two other conclusions of the theorem holds.

As in the proof of Proposition 3.1 we work on the h-torus T_0. This is justified by the observation that we can always find a collection of geodesics on $\tilde{T}$ that plays the same role with respect to $\tilde{\gamma}$ as $\mathscr{L}_h$. The intersection properties are crucial to the argument whereas the linearity is a pleasant simplification.

The idea of the argument is to study a lift $\tilde{\gamma}$ of the geodesic γ to $\tilde{T}$. As in the previous argument we see that $\tilde{\gamma}$ contains a piece that looks like a bump around some lattice points. The argument consists mainly of a case by case analysis of what goes on nearby the bump. The same reasoning that gave the bigon in Proposition 3.1 will be at work here.

It should be observed that the hypotheses of Theorem 3.1 and our normalization to T_0 preclude the possibility that a lift $\tilde{\gamma}$ will either terminate at a lattice point or spiral into a closed curve that bounds a disc about a single lattice point.

Choose a lift $\tilde{\gamma}$ of γ to $\tilde{T}$. Since γ is not simple there is some $g \in \Gamma$ for which $\tilde{\gamma}$ and $g(\tilde{\gamma})$ intersect transversely in a point p'. In other words there are points p and p' on $\tilde{\gamma}$ with $g(p)=p'$. We may normalize the covering so that $g(z)=z-m$ for some nonnegative integer m.

Let a be the linear horizontal segment joining p to p' and lying in the closure of the region Ω between the horizontal lines l_k and l_{k+1}. $\tilde{\gamma}$ must meet one of the lines l_k or l_{k+1} in two or more points. The alternatives are that $\tilde{\gamma}$ meets neither l_k nor l_{k+1}, $\tilde{\gamma}$ meets only one of l_k or l_{k+1} exactly once, or $\tilde{\gamma}$ contains a single arc joining l_k to l_{k+1}. In all three instances it is clear that the part of $\tilde{\gamma}$ in Ω is mapped disjointly from itself by every element of the group Γ.

Thus we may choose a lift $\tilde{\gamma}$ of γ which meets l_0 at least twice. Furthermore, this may be done as in the proof of Proposition 3.1 so that there is an arc c_0 of $\tilde{\gamma}$ (the bump) and an interval $a_0=[x,y]$ on l_0 bounding a region either above or below l_0 which is disjoint from $\mathbf{Z}[i]$ (see Figure 3.1). Evidently, for the intersection of c_0 with a_0 to be geometric a_0 must contain a lattice point.

The complement of the arc c_0 in $\tilde{\gamma}$ consists of a finite arc $\tilde{\gamma}_0$ and an infinite ray $\tilde{\gamma}_1$.

The following construction will be used to produce a simple closed geodesic that bounds a bigon with $\tilde{\gamma}$.

Suppose n_0 and n_1 are lattice points on the lines l_j and l_k respectively with $j \neq k$. There exist points p_0 lying between n_0-1 and n_0 on l_j and p_1 lying between n_1 and

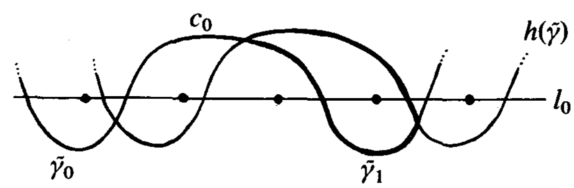

Fig. 3.3 (a)

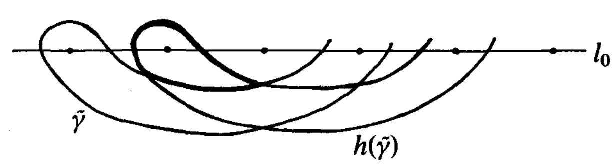

Fig. 3.3 (b)

n_1+1 on l_k so that the line l containing p_0 and p_1 has rational slope and is disjoint from $\mathbf{Z}[i]$. l projects to a simple closed noncontractible curve λ on T_0. Thus, there is a simple closed geodesic λ^* freely homotopic to λ. Lift the free homotopy between λ and λ^* to $\tilde{T}$ to produce a geodesic l^* freely homotopic to l on $\tilde{T}$ and covering λ^* on T_0. In general we let $l^*(n_0, n_1)$ denote a geodesic constructed in this way.

Six main cases are to be considered.

Case 1. l_0 is the only line in $\mathscr{L}_k$ that $\tilde{\gamma}$ intersects. Let $h(z)=z+1$. the geodesic $\tilde{\gamma}$ and its translate $h(\tilde{\gamma})$ will always together bound a bigon about one of the lattice points. To see this we consider two subcases.

First suppose that $\tilde{\gamma}$ contains a subarc consisting of three consecutive bumps. Without loss of generality the situation reduces to one of those illustrated in Figures 3.3 (a) and (b) where we have labeled the center bump c_0 and assume that the infinite ray $\tilde{\gamma}_1$ in the complement of c_0 on $\tilde{\gamma}$ intersects c_0 in the endpoint y. Let n be the smallest integer that is larger than y. The existence of a bigon about n bounded by $\tilde{\gamma}$ and $h(\tilde{\gamma})$ follows as in the proof of Proposition 3.1. The analysis given there shows that the proper intersections must occur.

Now we suppose that $\tilde{\gamma}$ does not contain three consecutive bumps. Thus the end of $\tilde{\gamma}$ must be asymptotic to one of the geodesics $\{\operatorname{Im} z=\frac{1}{2}\}$ or $\{\operatorname{Im} z=-\frac{1}{2}\}$. We suppose it is the latter. $\tilde{\gamma}$ must then contain a bump c_0 above l_0 directly preceding an infinite ray $\tilde{\gamma}_1\subset\tilde{\gamma}$ which is asymptotic to $\{\operatorname{Im} z=-\frac{1}{2}\}$ and meets l_0 only at its endpoint y. Let n again be the smallest integer larger than y. Then $\tilde{\gamma}$ and $h(\tilde{\gamma})$ must together bound a bigon about n with one ideal vertex. This follows easily if one keeps in mind that for some $g(z)=z-m$, $\tilde{\gamma}\cap g(\tilde{\gamma})\neq\emptyset$ (see Figures 3.3 (c) and (d)).

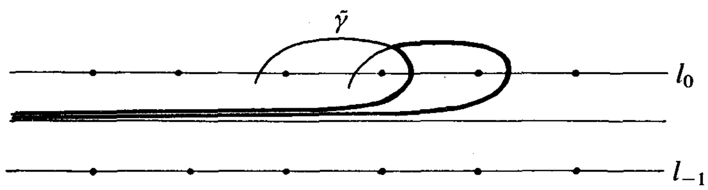

Fig. 3.3 (c)

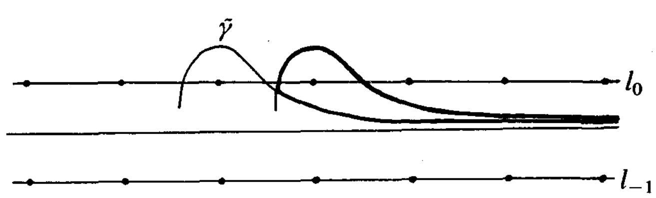

Fig. 3.3 (d)

For the remainder of the argument we suppose that $\tilde{\gamma}$ meets l_0 and another line l_{-n} for some nonnegative integer n.

Case 2. $\tilde{\gamma}$ contains two or more bumps lying on different lines. We suppose that one of the bumps c_0 lies above l_0 and is joined to a second bump c_1 lying below l_{-n} by an arc d which meets each line l_{-k}, with $0 \leq k \leq n$, in exactly one point. c_1 bounds an arc a_1 on l_{-n} with endpoints x' and y' where $\operatorname{Re} x' < \operatorname{Re} y'$. We will assume that d meets a_1 at the point x', and that the infinite ray in the complement of c_1 on $\tilde{\gamma}$ meets a_1 at y' (see Figure 3.4 (a)). This precisely describes the configuration in case 2 up to reflections in horizontal or vertical lines and translations.

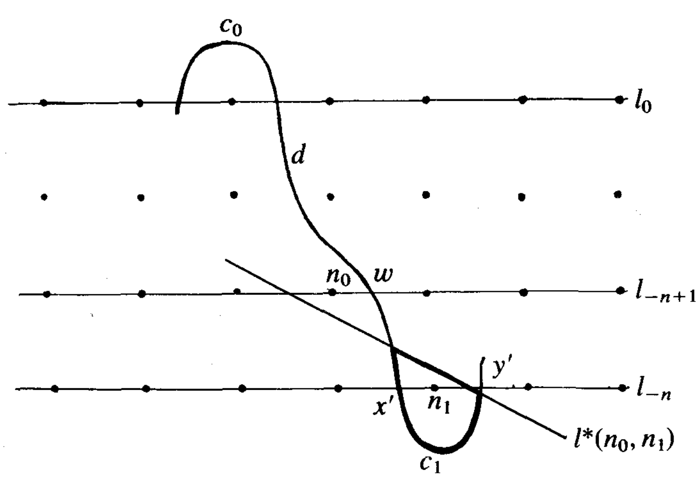

Fig. 3.4 (a)

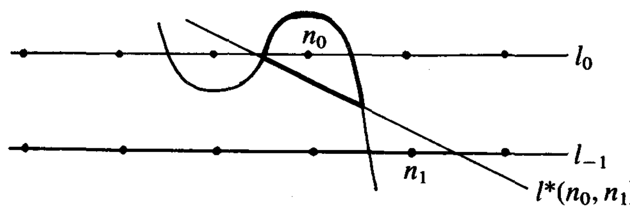

Fig. 3.4 (b)

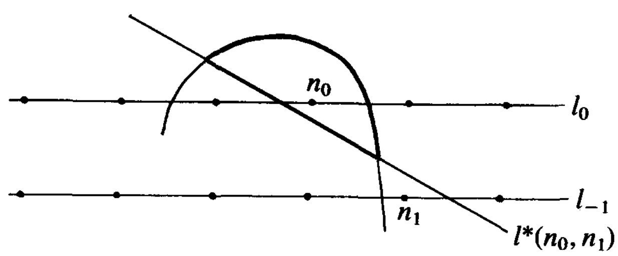

Fig. 3.4 (c)

Let w be the point at which d intersects l_{-n+1}. Let n_0 be the lattice point on l_{-n+1} which has maximal real part among all lattice points n with $\mathrm{Re}\, n < \mathrm{Re}\, w$. Let n_1 be the lattice point on a_1 with smallest real part. A geodesic $l^*(n_0, n_1)$ then bounds a bigon about n_1 along with $\tilde{\gamma}$. The intersection of l^* with d is clear. To see the second intersection of l^* with one of c_1 or the infinite ray meeting c_1 at y' one follows the analysis of Proposition 3.1.

The resolutions of the following two cases are so similar to the previous case that we will restrict ourself to providing illustrations which clearly indicate the appropriately normalized configurations and the location of points n_0 and n_1 so that $l^*(n_0, n_1)$ bounds a bigon with $\tilde{\gamma}$.

Case 3. There are exactly two bumps in $\tilde{\gamma}$ both of which lie on l_0 and bound disjoint intervals (see Figure 3.4 (b)).

Case 4. There is exactly one bump c_0 in $\tilde{\gamma}$ so that the arc a_0 sharing its endpoints contains two or more lattice points (see Figure 3.4 (c)).

Case 5. $\tilde{\gamma}$ contains only a single bump c_0 above l_0. Suppose that the infinite ray $\tilde{\gamma}_1$ in the complement of c_0 in $\tilde{\gamma}$ meets c_0 in the endpoint y. Then the finite arc $\tilde{\gamma}_0$ meets c_0 at the endpoint x. First suppose that $\tilde{\gamma}_0$ meets l_{-1} in a point z. Let n_0 be the lattice point on l_{-1} with maximal real part among all lattice points n on l_{-1} with $\mathrm{Re}\, n < \mathrm{Re}\, z$.

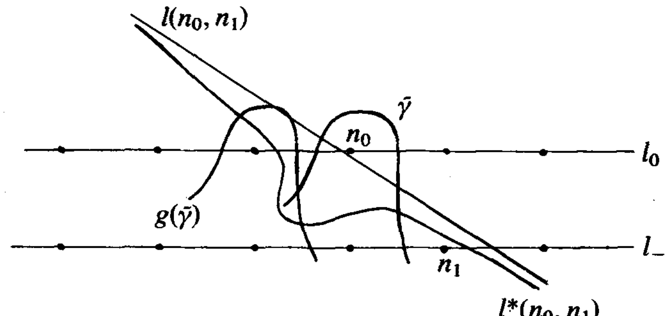

Fig. 3.4 (d)

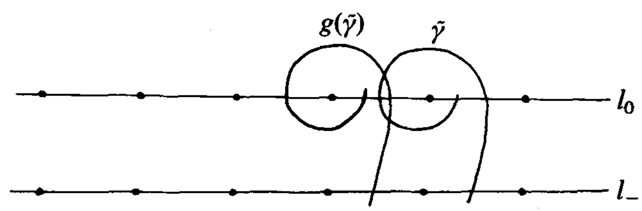

Fig. 3.4 (e)

Let n_1 be the smallest integer larger than x. Then a geodesic $l^*(n_0, n_1)$ bounds a bigon about n_1 along with $\tilde{\gamma}$.

Finally we suppose that except for its endpoint x, $\tilde{\gamma}_0$ lies entirely inbetween l_0 and l_{-1}. Since we have already considered cases 1, 3 and 4 we may suppose that $\tilde{\gamma}_1$ meets l_{-1} in a single point w and a_0 contains a single lattice point n_0. Let n_1 be the lattice point on l_{-1} with minimal real part among all lattice points n on l_{-1} with $\mathrm{Re}\, n > \mathrm{Re}\, w$.

In this case it is not immediately clear that $l^*(n_0, n_1)$ will meet the arc $\tilde{\gamma}_0 \cup c_0$ to produce a bigon. We need to make use of the fact that there is a transformation $g \in \Gamma$ with $\tilde{\gamma} \cap g(\tilde{\gamma}) \neq \emptyset$. Under the circumstances this implies that there are points $p' \in \tilde{\gamma}_0 \cup c_0$ and $p \in \tilde{\gamma}_1 \cup c_0$ with $g(p) = p'$. Moreover, $p \neq p'$ for otherwise $\tilde{\gamma}$ will bound a monogon about n_0. Without loss of generality we suppose that $p \in \tilde{\gamma}_1$ and $p' \in \tilde{\gamma}_0$. If l^* were to miss $\tilde{\gamma}_0 \cup c_0$ and thus avoid bounding a bigon, this would force there to be nongeometric intersections between l^* and $g(\tilde{\gamma}_1 \cup c_0)$ (see Figure 3.4 (d)), which cannot occur.

Case 6. There are exactly two concentric bumps in $\tilde{\gamma}$ both of which lie on l_0. This configuration cannot occur. To see this observe that, as in the last part of case 5, the intersection produced by $g(z) = z - m$ would occur between the arcs $\tilde{\gamma}_0 \cup C_0$ and $\tilde{\gamma}_1 \cup C_0$. Such an intersection is necessarily nongeometric (see Figure 3.4 (e)). Q.E.D.

In the case of closed geodesics the following theorem was originally asserted by Beardon, Lehner, and Sheingorn.

THEOREM 3.2. *Let γ be a self intersecting compactly supported geodesic on a* 4*x* *sphere M, and suppose that γ does not spiral around a boundary geodesic. Then γ bounds a monogon on M.*

Proof. Let

$$\Gamma(3)=\left\{\begin{pmatrix} a & b \\ c & d\end{pmatrix}\equiv\pm\begin{pmatrix} 1 & 0 \\ 0 & 1\end{pmatrix} \mod 3\right\}\subset SL(2,\mathbf{Z}),$$

which we view as a subgroup of $\text{Möb}_{\mathbf{R}}$. The quotient $\mathbf{H}/\Gamma(3)$ is a $4x$ sphere $M_{\mathbf{Z}}$. As in the proof of Theorem 3.1 it will suffice to prove the theorem on $M_{\mathbf{Z}}$.

Let Γ' be the subgroup of $\text{Möb}_{\mathbf{R}}$ with generators

$$\begin{pmatrix} 2 & 1 \\ 1 & 1\end{pmatrix},\begin{pmatrix} 1 & 1 \\ 1 & 2\end{pmatrix}\in SL(2,\mathbf{Z}).$$

The quotient $\mathbf{H}/\Gamma'$ is an h-torus $T_{\mathbf{Z}}$. It is easy to show that all parabolic Möbius transformations in Γ' are contained in $\Gamma(3)$. Both groups are finite index normal subgroups of $SL(2,\mathbf{Z})$.

We shall make use of an observation made by C. Series [20] that elementary geometric arguments may be employed to show that a geodesic in $\mathbf{H}$ projects to a simple closed geodesic on $M_{\mathbf{Z}}$ if and only if it projects to a simple closed geodesic on $T_{\mathbf{Z}}$. Sheingorn also has a proof of this fact which unfortunately depends upon the conclusion of the theorem we are proving. It follows immediately from this observation that a geodesic in $\mathbf{H}$ projects to a geodesic in $\bar{S}(M_{\mathbf{Z}})$ if and only if it projects to a geodesic in $\bar{S}(T_{\mathbf{Z}})$.

If γ is a self intersecting geodesic on $M_{\mathbf{Z}}$ or $T_{\mathbf{Z}}$, we may classify the type of intersection as follows: Let p be a point at which γ intersects itself. There is a lift $\tilde{p}$ of p to $\mathbf{H}$, and lifts $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ of γ to $\mathbf{H}$ meeting at $\tilde{p}$. Let R be the set of elements in the covering group mapping $\tilde{\gamma}_1$ to $\tilde{\gamma}_2$. We call p a parabolic intersection if R contains a parabolic Möbius transformation, otherwise p is a hyperbolic intersection. It is well known that an intersection p is parabolic if and only if it is the vertex of a geodesic arc in γ bounding a monogon.

First suppose that γ is a closed self intersecting geodesic on $M_{\mathbf{Z}}$. Choose a geodesic $\tilde{\gamma}_1$ covering γ in $\mathbf{H}$. Since γ is not simple $\tilde{\gamma}_1$ projects to a closed geodesic γ' on $T_{\mathbf{Z}}$ which cannot be simple. By Proposition 3.1 γ' bounds either a monogon or a bigon on $T_{\mathbf{Z}}$.

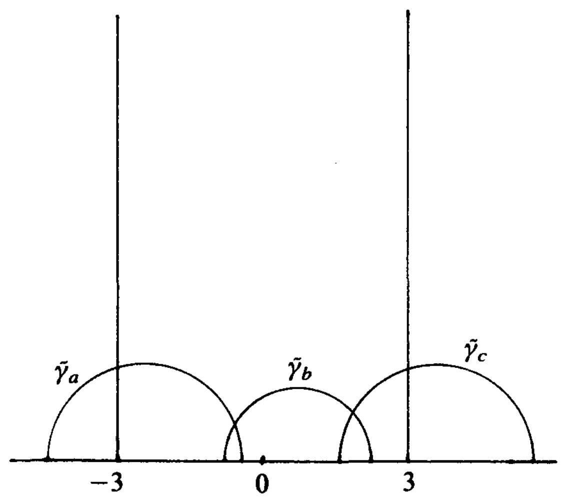

Fig. 3.5

If γ' bounds a monogon then it contains a parabolic self intersection and there must be a parabolic transformation $g \in \Gamma'$ with $g(\tilde{\gamma}_1) \cap \tilde{\gamma}_1 \neq \emptyset$. It follows that γ contains a parabolic intersection, and hence bounds a monogon on $M_{\mathbf{Z}}$.

We now consider the case where γ' bounds a bigon. The parabolic subgroup of Γ' stabilizing infinitely, $\mathrm{Stab}(\infty)$, is generated by $h(z)=z+6$. Consider the region W in $\mathbf{H}$ consisting of all z in $\mathbf{H}$ with $-3 \leq \mathrm{Re}\, z \leq 3$. W is a fundamental domain for the action of $\mathrm{Stab}(\infty)$ on $\mathbf{H}$.

Lift the bigon in γ' to $\mathbf{H}$ in such a way that it separates infinity from the real axis in W. This will result in three lifts $\tilde{\gamma}_a$, $\tilde{\gamma}_b$, $\tilde{\gamma}_c$ of γ' to $\mathbf{H}$ as illustrated in Figure 3.5. The transformation $h(z)$ maps $\tilde{\gamma}_a$ to $\tilde{\gamma}_c$.

An easy calculation reveals that one of the geodesics, say $\tilde{\gamma}_a$, lies on a circle of diameter greater than three.

The stabilizer of infinity in $\Gamma(3)$ is generated by the transformation $k(z)=z+3$. Then $k(\tilde{\gamma}_a) \cap \tilde{\gamma}_a \neq \emptyset$, and so $\tilde{\gamma}_a$ will project to a geodesic γ_a on $M_{\mathbf{Z}}$ with a parabolic self intersection, in other words, γ_a bounds a monogon on $M_{\mathbf{Z}}$.

Since $\Gamma(3)$ is normal in $SL(2,\mathbf{Z})$ each $g \in SL(2,\mathbf{Z})$ induces an isometric self map of $M_{\mathbf{Z}}$. Let g be the element of Γ' mapping $\tilde{\gamma}_a$ to $\tilde{\gamma}_1$. Such a transformation exists since both $\tilde{\gamma}_a$ and $\tilde{\gamma}_1$ cover γ'. Then g induces an isometry of $M_{\mathbf{Z}}$ which maps γ_a to γ. Thus we see that γ bounds a monogon.

Suppose now that γ is an open geodesic on $M_{\mathbf{Z}}$ with self intersections. As above lift γ to a geodesic $\tilde{\gamma}_1$ in $\mathbf{H}$, and project $\tilde{\gamma}_1$ down to a geodesic γ' on $T_{\mathbf{Z}}$. If γ' has self intersections then by Theorem 3.1 it must either bound a monogon, bound a bigon, or bound a bigon with the assistance of a simple closed geodesic α. The first two cases have already been dealt with. In the third case we can find three lifts of α and γ' in $\mathbf{H}$ bounding a neighborhood of infinity in W.

By Theorem 2.4 the largest horocycle $\{\operatorname{Im} z>c\}$ in W disjoint from a lift of α has area larger than four. It follows that any lift of α must lie on a circle of diameter less than three. One of the lifts $\tilde{\gamma}'$ of γ' bounding the neighborhood of infinity in W must therefore lie on a circle of diameter greater than three. Then either $\mathfrak{k}(\tilde{\gamma}')\cap\tilde{\gamma}'\neq\emptyset$ or $\mathfrak{k}^{-1}(\tilde{\gamma}')\cap\tilde{\gamma}'\neq\emptyset$. The earlier arguments now apply to give the monogon.

It is possible that the geodesic γ' on $T_{\mathbf{Z}}$ is simple. If this is the case then, since γ does not belong to $\bar{S}(M_{\mathbf{Z}})$, γ' does not belong to $\bar{S}(T_{\mathbf{Z}})$. From Proposition 4.3 of the next section we may deduce that if γ' is simple then it either bounds an ideal monogon on $T_{\mathbf{Z}}$ or it bounds a bigon along with a simple closed geodesic α.

The later possibility has already been considered so we will suppose γ' bounds an ideal monogon on $T_{\mathbf{Z}}$. Then γ' lifts to a geodesic $\tilde{\gamma}'$ in $\mathbf{H}$ which shares an endpoint with its translate $h(\tilde{\gamma}')$. It follows that $\tilde{\gamma}'$ lies on a circle of radius three, and hence that $\tilde{\gamma}'\cap\mathfrak{k}(\tilde{\gamma}')\neq\emptyset$. Again, earlier arguments demonstrate the existence of a monogon. Q.E.D.

§3.1. We are now prepared to argue the assertions of Theorems 1–4 concerned with self intersecting geodesics. These comprise the vast majority of all geodesics on a surface and in particular of those not in the set $\bar{S}$. After having arrived at the topological characterization of Theorems 3.1 and 3.2 the following proposition will be sufficient to complete this part of the argument.

PROPOSITION 3.2. *Let N be one of the surfaces T or M in Theorems* 1.3 *to* 1.6. *Let α be a geodesic on the surface N. Define $A(\alpha)$ with respect to a puncture p on N and $D(\alpha)$ with respect to an infinite noncompact end e with its associated boundary geodesic γ.*

(1) *If α bounds a monogon about p then $A(\alpha)\leq 2$ with equality occurring when α bounds only ideal monogons about p.*

(2) *If α bounds no monogon about p but does bound a bigon about p, either alone or along with a simple closed geodesic, then $2\leq A(\alpha)<4$.*

(3) *If α bounds a monogon about e then* $\cosh D(\alpha)\leq\coth\frac{1}{2}L(\gamma)$, *with equality occurring when α bounds only ideal monogons about e.*

(4) *If α bounds no monogon about e but does bound a bigon about e, either alone or along with a simple closed geodesic, then*

$$\coth\frac{L(\gamma)}{2} \leqslant \cosh D(\alpha) < \coth\frac{L(\gamma)}{4}.$$

It is worth noting that we are not aware of any examples to show equality can occur in statements (2) and (4) of the proposition. Nevertheless, it is clear from the proof that this would necessarily be a limiting case; that is, the distance from α to γ could not be realized by a geodesic arc.

Proof. Let G be a Fuchsian group representing the surface N. By conjugating in $\text{Möb}_{\mathbf{R}}$ we may normalize G so that the stabilizer of infinity in G is generated by the element $g(z)=z+1$. We may further suppose that g is a primitive parabolic representing the puncture p.

(1) If α bounds a monogon about p then there is a lift $\tilde{\alpha}_1$ of α which intersects its translate $\tilde{\alpha}_2=g(\tilde{\alpha}_1)$ either in $\mathbf{H}$ or at an endpoint. Denote the endpoints of $\tilde{\alpha}_1$ by ξ_1 and ξ_2 with $\xi_1<\xi_2$. Then we have $\xi_1<\xi_1+1\leqslant\xi_2$. It follows that $\tilde{\alpha}_1$ lies on a circle of radius $r\geqslant\frac{1}{2}$. Therefore, the lift of the horocycle about p of area $A(\alpha)$ must cover a neighborhood of infinity above height $\frac{1}{2}$. A fundamental domain for the horocycle above $\operatorname{Im} z=\frac{1}{2}$ is $F=\{z|0\leqslant\operatorname{Re} z\leqslant 1, \operatorname{Im} z\geqslant\frac{1}{2}\}$. Thus

$$A(\alpha)\leqslant \iint_F \frac{dx\,dy}{y^2}=2.$$

(2) To get the lower bound at 2 we notice that if $A(\alpha)<2$ then there is a lift $\tilde{\alpha}$ of α which enters the horocyclic neighborhood of infinity above the line $\operatorname{Im} z=\frac{1}{2}$. This geodesic $\tilde{\alpha}$ lies on a circle of radius greater than $\frac{1}{2}$. Consequently, $g(\tilde{\alpha})\cap\tilde{\alpha}\neq\emptyset$ or equivalently, α bounds a monogon about the puncture p.

If α bounds a bigon on N about p then there are lifts $\tilde{\alpha}_1$, $\tilde{\alpha}_2$, and $\tilde{\alpha}_3$ of α with $g(\tilde{\alpha}_1)=\tilde{\alpha}_3$, $\tilde{\alpha}_1\cap\tilde{\alpha}_2\neq\emptyset$, and $\tilde{\alpha}_2\cap\tilde{\alpha}_3\neq\emptyset$. When there is an ideal vertex one of the intersections will occur at an endpoint. It follows, as above, that either $\tilde{\alpha}_1$ or $\tilde{\alpha}_2$ lies on a circle with radius larger than $\frac{1}{4}$, and hence $A(\alpha)<4$.

Now suppose that α bounds a bigon along with a simple closed geodesic β. There is no loss of generality in assuming that N is an h-torus T. Then there are lifts $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ of α with $g(\tilde{\alpha}_1)=\tilde{\alpha}_2$, and a lift $\tilde{\beta}$ of β with $\tilde{\beta}\cap\tilde{\alpha}_1\neq\emptyset$ and $\tilde{\beta}\cap\tilde{\alpha}_2\neq\emptyset$. Again, if there is an ideal vertex then one of the intersections will occur at an endpoint. We may conclude, as before, that either $\tilde{\alpha}_1$ or $\tilde{\beta}$ lies on a circle of radius greater than $\frac{1}{4}$. By Theorem 2.4 $A(\beta)>4$ showing that $\tilde{\alpha}_1$ must have the larger radius. We conclude that $A(\alpha)<4$.

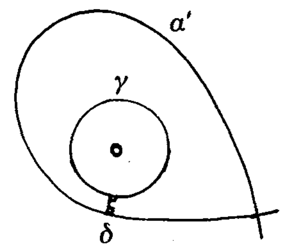

Fig. 3.6

(3) Let α' be a segment of α bounding a monogon about e. Choose a minimal length geodesic arc δ joining α' to the boundary geodesic γ. Since δ is locally length minimizing among all arcs joining γ and α it is a common perpendicular.

On N the arcs α', δ, and γ bound a simply connected region which lifts to a geodesic pentagon P in $\mathbf{H}$ (see Figure 3.6). From Theorem 2.1 we arrive at the identity

$$\cosh^2 L(\delta)+\cos\varphi = \sinh^2 L(\delta)\cosh L(\gamma).$$

Substituting for $\cos\varphi$ gives

$$\sinh^2 L(\delta)\cosh L(\gamma) \leqslant \cosh^2 L(\delta)+1$$

$$(\cosh^2 L(\delta)-1)\cosh L(\gamma) \leqslant \cosh^2 L(\delta)+1$$

$$\cosh L(\delta) \leqslant \left[\frac{\cosh L(\gamma)+1}{\cosh L(\gamma)-1}\right]^{1/2}$$

or

$$\cosh L(\delta) \leqslant \coth\frac{L(\gamma)}{2}.$$

Since $D(\alpha)<L(\delta)$ this shows that

$$\cosh D(\alpha) \leqslant \coth\frac{L(\gamma)}{2}.$$

It is clear that equality results only when $D(\alpha)=L(\delta)$ and $\varphi=0$. This is the case where every monogon bounded by α containing p is an ideal monogon.

(4) We will argue the case where α bounds a bigon about e along with a simple closed geodesic β. The case where α bounds alone follows easily.

Choose a minimal length geodesic arc δ joining γ to the boundary of the bigon. As in (3) δ is a common perpendicular. The argument will demonstrate that

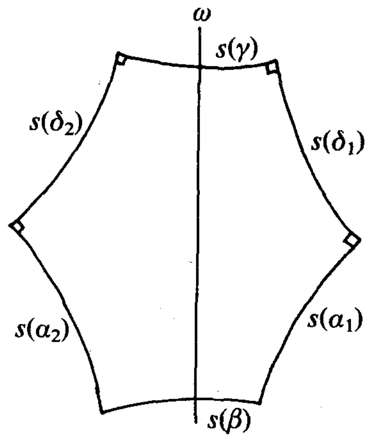

Fig. 3.7 (a)

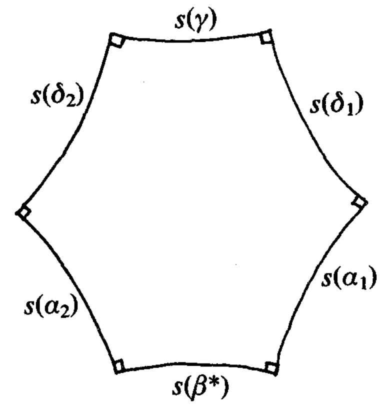

Fig. 3.7 (b)

$\cosh L(\delta)<\coth\frac{1}{4}L(\gamma)$. By Theorem 2.3 $\cosh D(\beta)>\coth\frac{1}{4}L(\gamma)$ so it will follow that δ joins γ to α. With this justification in mind there will be no loss of generality if we assume that δ joins γ to α. The bigon along with γ and δ bounds a simple connected region on N which lifts to a hexagon H in $\mathbf{H}$ (see Figure 3.7 (a)).

Let ω be the perpendicular bisector of the side $s(\gamma)$. Reflection in ω leaves $s(\gamma)$ invariant, exchanges the side $s(\delta_1)$ with $s(\delta_2)$, and exchanges the geodesic α_1 with α_2. Let β^* be the geodesic perpendicular to ω, and lying on the same side of γ in $\mathbf{H}$ as the hexagon at a distance $L(\delta)=d_i$, $i=1,2$ from γ. As δ is choosen to realize the minimal distance from γ to the boundary of the bigon it must be that either $\beta^*=\beta$ or else the point at which β^* meets ω lies in the interior of H. In either case β must meet one of the sides $s(\alpha_1)$ or $s(\alpha_2)$. Since reflection in ω fixes β^* and maps α_1 to α_2, β^* must meet both α_1 and α_2.

Define a new hexagon H^* by substituting β^* for β (see Figure 3.7 (b)). Then ω divides H^* into congruent pentagons. Applying Theorem 2.2 to one of these pentagons gives

$$\cosh^2 L(\delta)+\cos\varphi = \sinh^2 L(\delta)\cosh\frac{L(\gamma)}{2}.$$

Since we allow at most one ideal vertex $\varphi>0$. Manipulating as in (3) leads to the strict inequality

$$\cosh L(\delta) < \coth\frac{L(\gamma)}{4}.$$

As $D(\alpha)<L(\delta)$, $\cosh D(\alpha)<\coth\frac{1}{4}L(\gamma)$.

We need to show that if $\cosh D(\alpha)<\coth\frac{1}{2}L(\gamma)$ then α bounds a monogon. Choose geodesics $\tilde{\gamma}$ and $\tilde{\alpha}$ in $\mathbf{H}$, covering γ and α respectively, so that the common perpendicular δ joining $\tilde{\gamma}$ to $\tilde{\alpha}$ satisfies $L(\delta)<\coth\frac{1}{2}L(\gamma)$. Let g be a generator of the subgroup of the deck transformations that stabilizes $\tilde{\gamma}$. If α does not bound a monogon on N then $\tilde{\alpha}$ and $g(\tilde{\alpha})$ are disjoint. Let β be the common perpendicular joining α and $g(\alpha)$.

The geodesics $\tilde{\gamma}$, $\tilde{\alpha}$, $g(\tilde{\alpha})$, δ, $g(\delta)$ and β bound a right hexagon H. Let ω be the common perpendicular joining $\tilde{\gamma}$ to β. H is invariant under reflection in ω. Hence, ω divides H into two equivalent pentagons.

Applying Theorem 2.2 to one of the pentagons gives

$$\cosh\frac{L(\gamma)}{2}\cosh L(\alpha)=\sinh\frac{L(\gamma)}{2}\sinh L(\alpha)\cosh L(\delta).$$

Then

$$\coth\frac{L(\gamma)}{2}=\tanh L(\alpha)\cosh L(\delta)<\cosh L(\delta)<\coth\frac{L(\gamma)}{2}.$$

We conclude that $\tilde{\alpha}\cap g(\tilde{\alpha})\neq\emptyset$ producing a monogon on N. Q.E.D.

§ 4. Simple geodesics

The analysis of simple geodesics presented here is based on Thurston's work with geodesic laminations. Although a majority of the results are implicit in [22] if they are not explicitely stated there we will give proofs or else outline a proof when the train track theory is involved. Alternate approaches may be found in [13].

We refer to a geodesic belonging to a lamination as a *leaf* of the lamination. A leaf γ of a geodesic lamination $\mathscr{L}$ on a surface N is called *proper* if there is an arc on N which intersects $\mathscr{L}$ transversely in a single point on γ. We will call a geodesic γ *proper* if γ is a proper leaf of the lamination which is the closure of γ on N.

A sequence of geodesic laminations $\mathscr{L}_i$ on N converges to $\mathscr{L}$ if given any leaf γ in $\mathscr{L}$ there is a sequence of leaves $\gamma_i\in\mathscr{L}_i$ converging to γ.

PROPOSITION 4.1, [22]. (1) *Laminations with finitely many leaves are dense in* $\mathscr{GL}$.

(2) *A lamination contains only finitely many proper leaves and the remaining leaves are partitioned into finitely many minimal sublaminations.*

(3) *In a finite leaved lamination an end of a noncompact leaf either spirals around a closed geodesic or tends to a noncompact end of* N.

PROPOSITION 4.2. *Let α be a geodesic on a hyperbolic surface N belonging to a minimal lamination $\mathscr{L}$* in $\mathscr{GL}_0$. *Then there is a sequence of simple closed geodesics α_i converging to α.*

Outline of *Proof.* The lamination $\mathscr{L}$ supports a transverse measure. Every measured lamination is a limit of measured laminations containing only finitely many leaves. These finite leaved laminations $\mathscr{L}_i$ are constructed in terms of a train track approximation to $\mathscr{L}$. If $\mathscr{L}$ is supported on a compact subset of N then so is a train track approximation. Consequently, all leaves in the laminations $\mathscr{L}_i$ have compact support. A compactly supported finite leaved lamination must contain closed leaves. Some sequence of closed leaves α_i, with $\alpha_i \in \mathscr{L}_i$, will converge to a leaf β of $\mathscr{L}$. Since every leaf of a minimal lamination is dense in the lamination the α_i must also converge to α. Q.E.D.

PROPOSITION 4.3. *Let α be a proper compactly supported simple geodesic on an h-torus T. Suppose that α is not a boundary geodesic and that it does not spiral around a boundary geodesic. Then α satisfies one of the following:*

(1) *α bounds an ideal monogon.*
(2) *α bounds a bigon together with a simple closed geodesic.*
(3) *α lies in $\bar{S}$.*

Proof. The closure of α on T is a lamination $\mathscr{L}$. First we suppose that $\mathscr{L}$ does not contain a closed leaf. Let $\mathscr{L}^*$ be a minimal sublamination of $\mathscr{L}$. Clearly, α does not belong to $\mathscr{L}^*$. We will show that the complement of $\mathscr{L}^*$ on T is an ideal bigon (two ideal vertices). α must lie in this bigon and its ends tend towards ideal vertices. The ends cannot tend towards distinct vertices since then α would be one of the geodesics bounding the bigon. It follows that both ends of α tend towards the same ideal vertex. Then α must loop around the noncompact end producing a monogon.

We need to see why the complement of $\mathscr{L}^*$ is an ideal bigon. Certainly there is no simple closed non-boundary geodesic on $T \setminus \mathscr{L}^*$. For if there were one then by cutting along it we could infer that $\mathscr{L}^*$ is supported on a three holed sphere. Since the only simple geodesics on a three holed sphere are proper geodesics this is impossible.

It follows that the complement of $\mathscr{L}^*$ is planar, and moreover topologically equivalent to a punctured disc. One of the boundary components of $T \setminus \mathscr{L}^*$ is formed by geodesics in $\mathscr{L}^*$. In other words, $T \setminus \mathscr{L}^*$ is an ideal polygon containing the noncompact end.

For each ideal vertex v of this polygon we define a geodesic γ_v on $T \setminus \mathscr{L}^*$ with one

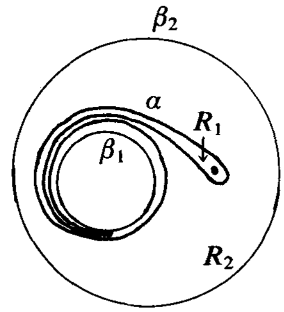

Fig. 4.1 (a)

end going to that vertex and the other end heading out to the noncompact end of T. If T has a boundary geodesic we may stipulate that γ_v is orthogonal to the boundary geodesic.

On a puncture h-torus T the complement of all the γ_v in $T\setminus\mathcal{L}^*$ is a collection of ideal triangles; one for each side of the polygon. An ideal triangle has area π. Since T has area 2π the polygon must be two sided.

If T has a boundary geodesic γ then each of the bounded regions on $T\setminus\mathcal{L}^*$ in the complement γ and the γ_v is a quadrilateral of area π. Since the bounded region on T in the complement of γ has area 2π we conclude again that the polygon is two sided. Thus we have shown that if $\mathcal{L}$ does not contain a closed geodesic then $T\setminus\mathcal{L}^*$ is an ideal bigon.

Next suppose that the closure of α on T is a lamination containing a closed leaf β. Again, since the complement of β on T is topologically a three holed sphere, β must be the unique minimal sublamination in $\mathcal{L}$.

It is assumed that β has compact support on T, and that α does not spiral around the boundary geodesic, if one exists. By Proposition 4.1 both ends of α must spiral around β. Cut T open along the geodesic β. The resulting surface T^* is topologically a three holed sphere; two of these holes have geodesic boundary β_1 and β_2 corresponding to the cut along β, and the third is a noncompact end.

The geodesic α lies on T^*. Each end of α must spiral around one of the geodesic boundaries β_1 or β_2. There are three cases to consider:

Case 1. Both ends of α spiral around the same geodesic which without loss of generality we suppose to be β_1. Since T^* is a planar surface there are two regions R_1 and R_2 in the complement of α on T^* (see Figure 4.1 (a)). If the region R_2, which has β_2 along its boundary, contains the noncompact end then R_1 is simply connected. A

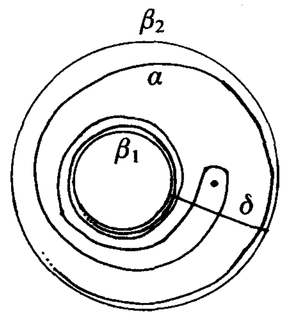

Fig. 4.1 (b)

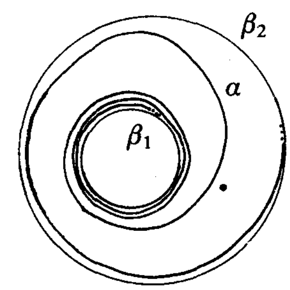

Fig. 4.1 (c)

simple geodesic cannot bound a simply connected region so the noncompact end must be in R_1. Therefore R_1 is a monogon.

Case 2. The two ends of α spiral around different bounding geodesics on T^* and both ultimately travel in a clockwise direction (with respect to an embedding of T^* in $\mathbf{C}$) (see Figure 4.1 (b)). There is a simple closed geodesic δ on T which on T^* becomes a geodesic arc joining β_1 to β_2 as illustrated. The region on T^* in the complement of the geodesics α and δ which contains the noncompact end is a bigon.

Case 3. The two ends of α spiral around different bounding geodesics on T^*; one end spiraling clockwise and the other counterclockwise (see Figure 4.1 (c)).

Let δ be as in case 2. The closed curves β and δ may be treated as generators for $\pi_1(T)$ where we select their intersection as the base point. Taking products in $\pi_1(T)$ define c_k to be the closed curve $\beta^k\delta$. Let γ_k be the unique closed geodesic in the free homotopy class of c_k. The γ_k are all simple and converge to α. Hence, in this case $\alpha \in \bar{S}$.

The convergence can be demonstrated by considering the symbolic dynamics of these geodesics. A fundamental domain may be chosen so that $\mathrm{Dyn}(\alpha)$ has a representative of the form

$$\sigma(n) = \begin{cases} b & n = 0 \\ a & \text{otherwise} \end{cases},$$

and $\mathrm{Dyn}(\gamma_k)$ has a representative of the form

$$\sigma_k(n) = \begin{cases} b & n = m(k+1) \text{ for some } m \in \mathbf{Z} \\ a & \text{otherwise} \end{cases}.$$

For any integer $N>0$, $k \geqslant N$, and $|n|<N$ $\sigma(n)=\sigma_k(n)$. If we interpret the symbolic dynamics in the context of Lemma 3.1 the convergence follows. Q.E.D.

PROPOSITION 4.4. *Let α be a proper compactly supported simple geodesic on a* 4x *sphere M. Suppose that α is not a boundary geodesic and that it does not spiral around a boundary geodesic. Then α lies in $\bar{S}$.*

Proof. As in Proposition 4.3 the closure of α on M is a geodesic lamination $\mathscr{L}$ containing a minimal sublamination $\mathscr{L}^*$. First suppose that $\mathscr{L}^*$ is a closed geodesic γ. The complement of γ consists of surfaces D_1 and D_2; each topologically a three holed sphere. α must lie on one of these surfaces; suppose it is D_1. γ is the only simple closed geodesic on D_1 which is not a boundary geodesic on M. Consequently, by Proposition 4.1, both ends of α must spiral around γ. As in case 1 in the proof of Proposition 4.3 α bounds a monogon on M. It is then an easy exercise to show that there is a sequence of simple closed curves on M converging to α.

Now consider the case where $\mathscr{L}$ does not contain a closed leaf. There must be at least two components in the complement of $\mathscr{L}^*$ on M, for otherwise $\mathscr{L}^*$ would lie in a simply connected subsurface. If there are exatly two complementary components then $\mathscr{L}^*$ lies in an annular subsurface of M, which cannot happen. If there are three complementary components then $\mathscr{L}^*$ lies in a three holed sphere on M, which is also impossible. We conclude that there are four components in $M \setminus \mathscr{L}^*$. Since each must contain a noncompact end of M they are all monogons. It is clearly impossible for the geodesic α to lie in any of these monogons. Therefore α belongs to $\bar{S}$. Q.E.D.

§4.1. It is now possible to tie all of the loose ends together and complete the proof of Theorems 1.3 to 1.6.

Let N be one of the surfaces T or M appearing in the hypotheses of Theorems 1.3 to 1.6. We define a subsurface N^* of N in the four cases:

(1.3) N^* is the set of points p on N whose distance $d(p,\gamma)$ from the boundary geodesic γ satisfies $\cosh d(p,\gamma)<\coth \frac{1}{4}L(\gamma)$.

(1.4) Let N_i^* be the set of points p on N whose distance $d(p,\gamma_i)$ from the boundary geodesics γ_i satisfies $\cosh d(p,\gamma_i)<\coth \frac{1}{4}L(\gamma)$. Set $N^*=\bigcup_{i=1}^4 N_i^*$.

(1.5) N^* is the open horocyclic neighborhood of the puncture having area 4.

(1.6) N^* is the union of the four open horocyclic neighborhoods of the punctures each having area 2.

Let $\overline{N^*}$ denote the closure of N^* in N. The proof will be complete if we show the following: A geodesic α belongs to $\bar{S}(N)$ if and only if α does not intersect N^*. If α belongs to $\bar{S}\setminus S$ then it intersects every open neighborhood of $\overline{N^*}$.

We will begin by considering geodesics α which do not lie in $\bar{S}$. In the case where α self intersects we may apply Proposition 3.2 along with the topological characterization given in Theorems 3.1 and 3.2 to conclude that α intersects N^*. If α does not self intersect then it must be simple. From Proposition 4.3 and 4.4 we infer that $N=T$ and α must either bound a monogon on N or bound a bigon on N along with a simple closed geodesic. Proposition 3.2 again allows us to conclude that α intersects N^*.

Let us now suppose that α belongs to $\bar{S}$. The results of (§ 2) provide exact values for $D(\alpha)$ and $A(\alpha)$ when α is a simple closed geodesic. It follows that $\alpha \in S$ is disjoint from N^*, and moreover, disjoint from $\overline{N^*}$.

By definition, for $\alpha \in \bar{S}\setminus S$ there exists a sequence of simple closed geodesics α_i converging to α. Since each one of the geodesics α_i is disjoint from $\overline{N^*}$, α must lie outside of N^*. For any value $k \in \mathbf{R}$ it is easy to show that there can only be a finite number of closed geodesics on N with length less than k. Consequently, the sequence of lengths $L(\alpha_i)$ must diverge to infinity. It follows, by computing $\lim_{i\to\infty} D(\alpha_i)$ or $\lim_{i\to\infty} A(\alpha_i)$, that we can choose points p_i on α_i so that the distance from p_i to N^* goes to zero. Some subsequence of the p_i converge to a point p on the boundary of $\overline{N^*}$. p must lie on a simple geodesic β which is a limit of the geodesics α_i.

If α is a proper geodesic then the α_i may be chosen so that the α_i will only converge to α and the closed geodesic that α spirals around. Since the geodesic that α spirals around must be disjont from $\overline{N^*}$ we get $\alpha=\beta$.

The final case to consider is when α is not a proper geodesic. Then α and β are both leaves of a lamination $\mathscr{L}$ in $\mathscr{GL}_0$ which does not contain a closed leaf.

Let $\mathscr{L}^*$ be a minimal sublamination of $\mathscr{L}$. We can infer from Proposition 4.3 and 4.4 and their proofs that $\mathscr{L}=\mathscr{L}^*$. Each leaf of $\mathscr{L}$ is dense in the entire lamination. This means that we can find a sequence of points q_i on α approaching p on β. We conclude that α intersects every open neighborhood of $\overline{N^*}$. Q.E.D.

§ 5. Diophantine approximation

LEMMA 5.1. *Let N be a hyperbolic surface with only finite noncompact ends. Let $\tilde{\gamma}$ be a geodesic ray in* $\mathbf{H}$ *with endpoint p. Suppose $\tilde{\gamma}$ covers a simple geodesic ray γ on N. Then either p is the fixed point of a parabolic transformation in the covering group G or p is the endpoint of a lift of some geodesic in $\bar{S}$.*

Proof. Assume that p is not the fixed point of a parabolic transformation in G, and that $\tilde{\gamma}$ does not lie on a geodesic $\tilde{\alpha}$ which projects to a geodesic $\alpha \in \bar{S}$.

We will begin by showing that the closure of γ on N, $\bar{\gamma}$, contains a geodesic lamination $\mathscr{L}$. Consider the collection of all G-translates of $\tilde{\gamma}$ that cross a finite sided fundamental domain F. Since $\tilde{\gamma}$ does not terminate at a parabolic fixed point, infinitely many of these must cross a compact subset of F. It follows that there is a geodesic $\tilde{\alpha}$ crossing F so that the arc $\tilde{\alpha} \cap F$ is a limit of translate of $\tilde{\gamma}$ in F. In other words, there are transformations $g_i \in G$ so that $g_i(\tilde{\gamma}) \cap F$ converges to $\tilde{\alpha} \cap F$.

By [15] there is a subsequence g_i^* and points $x, y \in \hat{\mathbf{R}}$ so that $g_i^*(z)$ converges to x for all $z \in \mathbf{C} \setminus \{y\}$. Let p' be the endpoint at the compact end of $\tilde{\gamma}$ in $\mathbf{H}$. Then $\lim_{i\to\infty} g_i^*(p')=x$. Since $g_i^*(\tilde{\gamma})$ all meet a compact subset of F, $\lim_{i\to\infty} g_i^*(p)=w \neq x$, i.e., $w=y$. The points x and w must be the endpoints of $\tilde{\alpha}$.

We have shown that if a piece of a geodesic α on N lies in the closure of γ then all of α does. Since the ray γ was assumed not to lie on a geodesic in $\bar{S}$, $\bar{\gamma}-\gamma$ is a geodesic lamination $\mathscr{L}$ containing a minimal sublamination $\mathscr{L}^*$.

Suppose $\mathscr{L}^*$ is a simple closed geodesic α. Let $\tilde{\alpha}$ be a lift of α with endpoints x and w. Without loss of generality there is a sequence of transformations $g_i \in G$ with $\lim_{i\to\infty} g_i(p)=w$ and $\lim_{i\to\infty} g_i(p')=x$. Let A be a hyperbolic transformation in G which fixes the geodesic $\tilde{\alpha}$ and has x as an attractive fixed point. It follows easily that if $p \neq w$ then for all but a finite number of the g_i we get $g_i(\tilde{\gamma}) \cap \tilde{\gamma} \neq \emptyset$. Hence, $\tilde{\gamma}$ must share an endpoint with $\tilde{\alpha}$.

Now we consider the remaining case where $\mathscr{L}$ contains no closed leaves. The connected components in $N \setminus \mathscr{L}^*$ are finite in number; each is a subsurface N_i of N with finitely many leaves of $\mathscr{L}^*$ along its boundary. The finiteness follows from an area argument similar to the one in the proof of Proposition 4.3 [22]. γ must lie on one of the surfaces N_i. We will suppose that γ lies on N_1.

Corresponding to each boundary component of N_1 there is a simple closed curve c_j which bounds a topological annulus along with the given boundary component (see Figure 5.1). The complement of all of the curves c_j on N_1 consists of some number of noncompact topological annuli and a single subsurface N_1^* bounded only by the curves c_j. Since it is assumed that γ does not head out to a puncture on N, $\gamma \cap \overline{N_1^*}$ must lie in a compact subsurface of $\overline{N_1^*}$. γ does not limit at any point on $\overline{N_1^*}$; therefore only a finite length segment of γ lies in $\overline{N_1^*}$. It follows that all but a finite length segment of γ lies in one of the annuli $\mathscr{A}$, and that the end of γ must be asymptotic to a geodesic α in $\mathscr{L}^*$ along the boundary of $\mathscr{A}$. Hence, there is a lift $\tilde{\alpha}$ of α with one endpoint at p. Q.E.D.

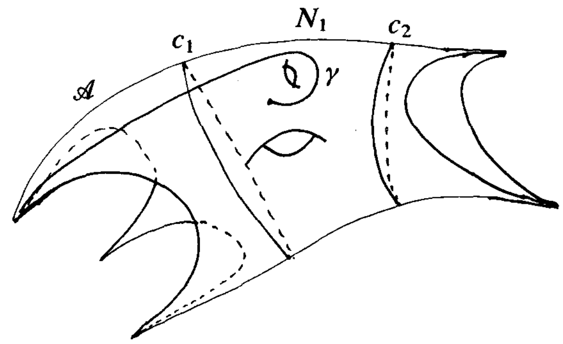

Fig. 5.1

A geodesic ray γ is said to be *essentially simple* if γ contains a simple ray. Lemma 5.1 implies that a geodesic ray γ is essentially simple only if there is a simple geodesic on N which shares an end with γ. We do not need to stipulate that this simple geodesic lies in $\bar{S}$ since the lemma also shows that an endpoint of a lift of any simple geodesic is either an endpoint of a lift of a simple geodesic in $\bar{S}$ or a parabolic fixed point. Every parabolic fixed point is certainly an endpoint of a lift of many simple geodesics.

It is necessary before proceeding with the proof of Theorems 1.1 and 1.2, to observe that Theorem 3.2 can be interpreted as a statement about self intersecting geodesic rays on a $4x$ sphere.

THEOREM 5.1. *Let γ be a self intersecting geodesic ray on a $4x$ sphere M which does not spiral around a boundary geodesic. Then either γ bounds a monogon on M or γ is asymptotic to the end of a geodesic in $\bar{S}(M)$.*

Proof. As in the proof of Theorem 3.2 it will suffice to work with the surface $M_{\mathbf{Z}}$. Also, refer back for the notation. We need only consider the case where γ is not a closed geodesic.

If the geodesic ray γ' on $T_{\mathbf{Z}}$ corresponding to γ has self intersections then the argument proceeds exactly as before using Theorem 3.1. If γ' does not have self intersections then as a consequence of Lemma 5.1 the endpoints of any lift of γ' to $\mathbf{H}$ is the endpoint of a geodesic in $\bar{S}(T_{\mathbf{Z}})$. It follows that the endpoint of any lift of γ to $\mathbf{H}$ is similarly the endpoint of a geodesic in $S(M_{\mathbf{Z}})$. The result follows. Q.E.D.

§5.1. *The Proof of Theorems* 1.1 *and* 1.2. For $x \in \mathbf{R}$ let $\tilde{l}(x)$ be the vertical ray $\tilde{l}(x)=\{z|\operatorname{Re} z=x,\ 0\leq \operatorname{Im} z\leq 1\}$ in $\mathbf{H}$. Denote by $l(x)$ the geodesic ray on N that is covered by $\tilde{l}(x)$.

Let N be one of the surfaces M or T appearing in the statements of Theorems 1.1 and 1.2. We define N_k^* as the union of open horocycles of area k on N. Let $\mathrm{Hor}(k)$ be the full preimage of N_k^* in $\mathbf{H}$. $\mathrm{Hor}(k)$ is the orbit of the horocycle $U_k=\{z|\mathrm{Im}\, z>t/k\}$ under the action of the appropriate group: G or H^*.

If

$$g=\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{with } c \neq 0$$

then $g(U_k)$ is a horocycle of radius $k/(2tc^2)$ which is tangent to the real axis at the point $g(\infty)=a/c$. It follows that $\tilde{l}(x)\cap g(U_k)\neq\emptyset$ if and only if $|x-g(\infty)|<k/(2tc^2)$.

Let A take the value 2 or 4 respectively depending on whether N is M or T. Let $A(\alpha)$ be defined for a simple closed geodesic α as in the conclusion of Theorems 1.5 and 1.6.

In this setting Theorems 1.1 and 1.2 have the following formulation: x is the endpoint of a lift $\tilde{\alpha}(x)$ of a simple closed geodesic $\alpha(x)$ on N if and only if for some $K>A$ $\tilde{l}(x)$ intersects only finitely many of the horocycles in $\mathrm{Hor}(k)$ for $k\leq K$. Let x be the endpoint of a lift of a simple closed geodesic $\alpha(x)$ on N; then $\tilde{l}(x)$ intersects infinitely many horocycles in $\mathrm{Hor}(k)$ only if $k\geq A(\alpha)$. Let x be the endpoint of a lift of a simple geodesic in $\bar{S}\setminus S$; then $l(x)$ intersects infinitely many horocycles in $\mathrm{Hor}(k)$ if and only if $k\geq A$.

We will prove the theorems in the above form. First consider a point x which is not the endpoint of a lift of a simple geodesic on N. By Lemma 5.1 the ray $l(x)$ on N is not essentially simple. Define subrays $\tilde{l}_n(x)=\{z|\mathrm{Re}=x,\ 0\leq\mathrm{Im}\, z\leq 1/n\}$ of $\tilde{l}(x)$ and their projections $l_n(x)$ on N. Each ray $l_n(x)$ self intersects and satisfies the hypotheses of Theorem 3.1 or 5.1. Applying Proposition 3.2 as in §4.1, we conclude that each ray $l_n(x)$ meets N_A^*.

Since x cannot be in the orbit of infinity, $\tilde{l}(x)$ will intersect a horocycle in $\mathrm{Hor}(k)$ for any $k>0$ on only a finite length arc. It follows that $\tilde{l}(x)$ must intersect infinitely many of the horocycles in $\mathrm{Hor}(A)$.

Now suppose that x is the endpoint of a geodesic $\tilde{\alpha}(x)$ which projects to a simple closed geodesic $\alpha(x)$ on N. If $k<A(\alpha(x))$ then by Theorems 1.5 and 1.6 there is a tubular neighborhood about the geodesic $\alpha(x)$ on N which is disjoint from N_k^*. Lift this neighborhood to a tubular neighborhood about the geodesic $\tilde{\alpha}(x)$ in $\mathbf{H}$. Since $\tilde{l}(x)$ is asymptotic to $\tilde{\alpha}(x)$ all but a finite length segment of $\tilde{l}(x)$ lies inside this neighborhood. From the discreteness of the group we can infer that there are only finitely many horocycles in $\mathrm{Hor}(k)$ which meet $\tilde{l}(x)$ and have radii larger than a given value. Thus, $\tilde{l}(x)$ can intersect only finitely many horocycles in $\mathrm{Hor}(k)$.

On the other hand, if $k>A$ then by Theorems 1.5 and 1.6 there is a point p on $\alpha(x)$ in the interior of N_k^*. Since $l(x)$ spirals around $\alpha(x)$ there is a sequence of points p_n on $l(x)$ converging to p. We may suppose that each point p_n lies inside N_k^*.

Lift the p_n to points $\tilde{p}_n$ on $\tilde{l}(x)$. The $\tilde{p}_n$ are all contained in $\operatorname{Hor}(k)$ and $\lim_{n\to\infty} \operatorname{Im} \tilde{p}_n=0$. Hence $\tilde{l}(x)$ meets infinitely many horocycles in $\operatorname{Hor}(k)$.

After the proof of Theorem 1.2 we observed that $N_{A(\alpha)}^*$ is tangent to $A(\alpha)$ at two points from opposite sides of α. It is then clear from the above that $\tilde{l}(x)$ meets infinitely many horocycles in $\operatorname{Hor}(A(\alpha(x))$.

We may conclude that $\tilde{l}(x)$ intersects infinitely many horocycles in $\operatorname{Hor}(k)$ only when $k \geqslant A(\alpha(x))$.

The last case to consider is when x is the endpoint of a geodesic $\alpha(x)$ which projects to an open simple geodesic on N. x is not in the orbit of infinity, so by Lemma 5.1 we may suppose that $\alpha(x)$ is a leaf of a minimal geodesic lamination $\mathscr{L}$. Then $\alpha(x) \in \bar{S} \setminus S$ and Theorems 1.5 and 1.6 imply that the largest horocycle on N disjoint from $\alpha(x)$ is N_A^*. The result then follows using the same arguments given above when $\alpha(x) \in S$. Q.E.D.

§ 6. Symbolic dynamics

Let G be a Fuchsian group representing an h-torus T. The fundamental domain F and the set $\mathscr{S}$ of sides of F are defined as in § 1.3. Let λ be a geodesic that crosses F and does not terminate at a parabolic fixed point of G or at a point which does not lie in the limit set of G on $\hat{\mathbf{R}}$. Choose a point p at which x meets a side of F.

We will define a function $\sigma: \mathbf{Z} \to \mathscr{S}$ that describes the sequence of G-translates of F encountered along λ. The G-translates of F cover λ and their sides divide λ into geodesic segments which we shall call F-segments. Endpoints of the F-segments on λ are naturally in one-to-one correspondence with $\mathbf{Z}$. The base point p is identified with 0. λ is divided into two geodesic rays λ^+ and λ^- eminating from p. An endpoint is associated with the integer n if there are $|n|$ F-segments lying between it and p; n is positive if the endpoint lies on λ^+ and negative otherwise.

Let λ_n denote the F-segment bounded by the nth and the $(n+1)$st endpoints. There is a unique $g \in G$ which maps λ_n into F. Define $\sigma(n)$ to be the side of F to which the nth endpoint is mapped by g.

Given a geodesic γ on T which does not terminate at a noncompact end it is easily seen that all lifts of γ meeting F determine equivalent functions (§ 1.3) σ independently of the choice of a base point or of an orientation. We call this equivalence class the *symbolic dynamics* of γ, and denote it $\operatorname{Dyn}(\gamma)$.

It is known [19] that a geodesic is uniquely determined by its symbolic dynamics. Also, if T has an infinite noncompact end then a function $\sigma: \mathbf{Z} \to \mathcal{S}$ represents the symbolic dynamics of a geodesic on T if and only if σ is freely reduced.

The sides of the fundamental domain F, project to a pair of disjoint simple geodesics on T. Let α and β respectively denote the projections of the two sides a and b of F. Observe that the ends of α and β are asymptotic to the noncompact end of T.

On the surface $T_0 = \tilde{T}/\Gamma$ (see § 3) let α_0 and β_0 respectively denote the projections of the geodesics in $\mathcal{L}_v$ and $\mathcal{L}_h$. By standard results from differential topology there exists a diffeomorphism $f: T_0 \to T$ with $f(\alpha_0) = \alpha$ and $f(\beta_0) = \beta$. Composing the covering map $\pi_0: \tilde{T} \to T_0$ with f gives a covering $\pi: \tilde{T} \to T$. The lines in $\mathcal{L}_v$ cover α and the lines in $\mathcal{L}_h$ cover β. We may lift the metric on T to $\tilde{T}$ so that π becomes a local isometry.

It is preferable to define the symbolic dynamics of a geodesic in terms of a lift to $\tilde{T}$. Let $\tilde{\gamma}$ be a geodesic on $\tilde{T}$ which does not terminate at a lattice point. Choose a base point p at which $\tilde{\gamma}$ intersects one of the lines in $\mathcal{L}_\#$. Also, choose an orientation for $\tilde{\gamma}$. As in the original definition of symbolic dynamics there is a natural correspondence between the point at which $\tilde{\gamma}$ intersects lines in $\mathcal{L}_\#$ and the integers. p is identified with 0 and the numbers increase in the direction of the orientation of $\tilde{\gamma}$. Define a function $\sigma: \mathbf{Z} \to \mathcal{S}$. $\sigma(n)$ takes the value a^ε when the nth intersection along $\tilde{\gamma}$ is with a vertical line l_v, and $\varepsilon = +1$ or -1 depending on whether the orientation vector on $\tilde{\gamma}$ at the intersection is directed into the half plane to the right side or to the left side of l_v. $\sigma(n)$ takes the value b^ε when the nth intersection along $\tilde{\gamma}$ is with a horizontal line l_h, and $\varepsilon = +1$ or -1 depending on whether the orientation of $\tilde{\gamma}$ is directed into the half plane above or below l_h.

The function $\sigma: \mathbf{Z} \to \mathcal{S}$ resulting from the above carries the same information as the function defined in $\mathbf{H}$ by a lift $\tilde{\tilde{\gamma}}$ of $\tilde{\gamma}$ with the choosen base point $\tilde{p}$ covering p. The only difference possible is that the values a, a^{-1} (b, b^{-1}) may be interchanged. This can be remedied by composing the covering map with an appropriate automorphism of the surface T. Thus we may suppose that the equivalence class of functions defined as above in terms of a lift of γ to $\tilde{T}$ is exactly $\mathrm{Dyn}(\gamma)$.

§ 6.1. Let $\tilde{l}(x, c)$ with $(x, c) \in \mathbf{R}^2$ be a line in $\tilde{T}$ with slope x passing through the point ic. The vertical line passing through c is written $\tilde{l}(\infty, c)$.

The above definitions may be applied to define a function $\sigma: \mathbf{Z} \to \mathcal{S}$ describing the sequence of intersections $\tilde{l}(x, c)$ realizes with the lines in $\mathcal{L}_\#$. Different choices of base point or orientation lead to equivalent functions. Also, the Γ-translates of $\tilde{l}(x, c)$ define equivalent functions.

Let $l(x,c)$ denote the projection of $\tilde{l}(x,c)$ to T. The equivalence class of functions $\sigma\colon \mathbf{Z}\to\mathscr{S}$ determined by lifts of $l(x,c)$ to $\tilde{T}$ is the *symbolic dynamics* of the curve $l(x,c)$; written $\mathrm{Dyn}(l(x,c))$. For example, the curves $l(\infty,c)$ and $l(0,c)$ have representative functions $\sigma_\infty(n)=b$, $\sigma_0(n)=a$ for all $n\in\mathbf{Z}$.

Since an oriented line l will cross all lines in $\mathscr{L}_v$ or $\mathscr{L}_h$ from the same direction we see that a symbol in $\mathscr{S}$ and its inverse cannot both appear in the range of a representative σ for $\mathrm{Dyn}(l)$. It follows that a representative σ for $\mathrm{Dyn}(l)$ will always be a freely reduced function.

LEMMA 6.1. $\mathrm{Dyn}(l(x,c))$ *is represented in condensed form by the function* $\Sigma(x,c)$.

Proof. We may suppose that $x\neq 0,\infty$. The translation invariance further allows us to assume that $0<c<1$.

Consider the finite group of orthogonal transformations generated by $\varphi_1(z)=\bar{z}$ and $\varphi_2(z)=e^{\pi i/2}z$. These transformations leave invariant the grid $\mathscr{L}_\#$ and induce automorphisms $\varphi_1^\#$ and $\varphi_2^\#$ of the set $\mathscr{S}$ given by:

$$\varphi_1^\#(a)=a,\quad \varphi_1^\#(a^{-1})=a^{-1},\quad \varphi_1^\#(b)=b^{-1},\quad \varphi_1^\#(b^{-1})=b;$$

$$\varphi_2^\#(a)=b,\quad \varphi_2^\#(a^{-1})=b^{-1},\quad \varphi_2^\#(b)=a^{-1},\quad \varphi_2^\#(b^{-1})=a.$$

If σ is a function representing $\mathrm{Dyn}(l(x,c))$ then the functions $\varphi_1^\#\circ\sigma$ and $\varphi_2^\#\circ\sigma$ represent $\mathrm{Dyn}(\varphi_1(l(x,c)))$ and $\mathrm{Dyn}(\varphi_2(l(x,c)))$ respectively. Since the translates of $\tilde{l}(x,c)$ for $1\leq x<\infty$ are all of the $\tilde{l}(x,c)$ with $x\neq 0$, ∞ it is no loss of generality if we restrict our attention to the case where $1\leq x<\infty$.

Orient $\tilde{l}(x,c)$ in the direction of increasing real part and choose c as a base point. A value $m\in\mathbf{Z}$ for which $\sigma(m)=a$ corresponds to an intersection of $\tilde{l}(x,c)$ with a vertical line $l_v(n)$ at the point $n+i(nx+c)$ for some $n\in\mathbf{Z}$. Observe that between the intersections $n+i(nx+c)$ and $n+1+i((n+1)x+c)$ of $\tilde{l}(x,c)$ with $l_v(n)$ and $l_v(n+1)$ the imaginary part of $\tilde{l}(x,c)$ changes by an increment of $x\geq 1$. Consequently, between any two intersections of $\tilde{l}(x,c)$ with lines in $\mathscr{L}_v$ there must be an intersection with a line in $\mathscr{L}_h$. In other words, if $\sigma(m)=a$ then $\sigma(m+1)=b$.

We may therefore write

$$\Sigma(n)=\begin{cases} a & n=2m \\ b^{u(m)} & n=2m+1 \end{cases}$$

for some function $u\colon\mathbf{Z}\to\mathbf{Z}^+$. The function $u(m)$ counts the number of intersections $\tilde{l}(x,c)$ realizes with lines in $\mathscr{L}_h$ between the intersections with $l_v(m)$ and $l_v(m+1)$. This is equal to the number of integers between $mx+c$ and $(m+1)x+c$, which is precisely $q(x, c)(m)$. Q.E.D.

§ **6.2.** *Proof of Theorem* 1.7. We have shown that symbolic dynamics $\mathrm{Dyn}(l(x, c))$ for lines in $\tilde{T}$ are represented by freely reduced functions $\sigma: \mathbf{Z} \to \mathcal{S}$. By [19] there is a geodesic, which we denote by $\lambda(x, c)$, having the same symbolic dynamics as $l(x, c)$. We will show that the lines $l(x, c)$ in fact describe the symbolic dynamics of all simple geodesics in $\tilde{S}(T)$ which belong to some minimal lamination.

As Birman and Series observe in [3] there must be lifts $\tilde{l}(x, c)$ and $\tilde{\lambda}(x, c)$ of $l(x, c)$ and $\lambda(x, c)$ to $\mathbf{H}$ which have the same endpoints. From this we may infer that $l(x, c)$ and $\lambda(x, c)$ are freely homotopic curves on T.

First let us suppose that $x \in \hat{\mathbf{Q}}$. We may write x as p/q where $p, q \in \mathbf{Z}$ have no common divisors. $\tilde{l}(x, c)$ lies entirely in $\tilde{T}$ if and only if $c \neq mx+k$ for $m, k \in \mathbf{Z}$. The subgroup $\mathrm{Stab}(\tilde{l}(x, c))$ of Γ leaving $\tilde{l}(x, c)$ invariant is generated by a translation $z \mapsto z+q+ip$. Since all translations in $\Gamma \setminus \mathrm{Stab}(\tilde{l}(x, c))$ map $\tilde{l}(x, c)$ disjointly away from itself $l(x, c)$ is a simple closed curve on T. Hence, $\lambda(x, c)$ is a simple closed geodesic.

For different numbers c and c' the lines $\tilde{l}(x, c)$ and $\tilde{l}(x, c')$ are both stabilized by the same subgroup of Γ. It follows that $\lambda(x, c)$ and $\lambda(x, c')$ are homologous curves on T. From Nielsen's paper [16] it can be deduced that there is a unique simple closed geodesic in each non-trivial primitive homology class on T. Therefore $\lambda(x, c) = \lambda(x, c')$. We may drop the reference to c and write $\lambda(x)$ when $x \in \hat{\mathbf{Q}}$.

It is clear that for $x, x' \in \hat{\mathbf{Q}}$, $\lambda(x) = \lambda(x')$ if and only if $x = x'$. Thus $\hat{\mathbf{Q}}$ parameterizes a class of simple closed geodesics on T. It turns out that every simple closed geodesic $\lambda \in S(T)$ is of the form $\lambda(x)$ for some $x \in \hat{\mathbf{Q}}$. To see this lift λ to a geodesic $\tilde{\lambda}$ on $\tilde{T}$. Since λ is nontrivial in homology $\mathrm{Stab}(\tilde{\lambda})$ is generated by a transformation of the form $z \mapsto z+q+ip$ for some $p, q \in \mathbf{Z}$. Therefore λ is homologous to $\lambda(p/q)$ and applying Nielsen we get $\lambda = \lambda(p/q)$. Then $S = \{\lambda(x) \mid x \in \hat{\mathbf{Q}}\}$ and it follows that $\mathrm{Dyn}(\lambda)$ is of the form $\Sigma(x, c)$ for some $x \in \hat{\mathbf{Q}}$ and for any choice of an irrational number c.

Now we consider what happens when x is irrational. The minimal compactly supported laminations which do not contain closed leaves are shown to be parameterized as above, by the irrational numbers. The leaves of a lamination $\mathcal{L}(x)$ and their symbolic dynamics may be described precisely by the lines $l(x, c)$.

For x irrational $\mathrm{Stab}(\tilde{l}(x, c))$ is trivial. Thus $l(x, c)$ is a simple open curve on T. Since $\mathrm{Dyn}(l(x, c))$ is distinct from $\mathrm{Dyn}(\lambda(p/q))$ for any $p/q \in \hat{\mathbf{Q}}$, $\lambda(x, c)$ must be a simple open geodesic on T. The line $l(x, c)$ can be mapped to a line $l(x, c')$ by an element of Γ if and only if $c' = c+kx+m$ for some integers k and m. Moreover, these are also necessary

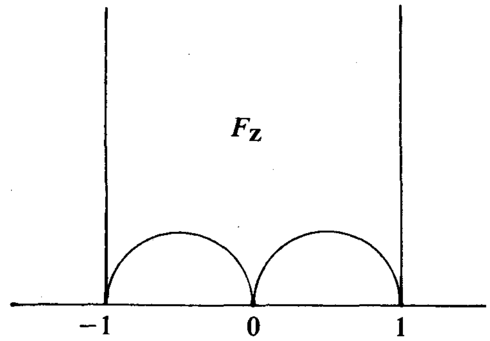

Fig. 6.1

and sufficient conditions for having $\mathrm{Dyn}\,(l(x,c))=\mathrm{Dyn}\,(l(x,c'))$. Consequently, there exists a continuum of disjoint simple geodesics $\lambda(x,c)$ for a fixed irrational number x.

It is well known that when x is irrational the Γ-translates of $\tilde{l}(x,c)$ intersect the imaginary axis in a dense set of points. It follows that for any number c' we can find a sequence of Γ-translates of $\tilde{l}(x,c)$ converging to $\tilde{l}(x,c')$. Applying Lemma 3.1 we may conclude that $\lambda(x,c)$ converges to $\lambda(x,c')$. Therefore the geodesics $\lambda(x,c)$ are all leaves of a minimal lamination $\mathscr{L}(x)$.

We must look more closely at what happens if we choose $c'=0$. The line $\tilde{l}(x,0)$ does not lie entirely in $\tilde{T}$ since it passes through the origin. Nonetheless we can find a sequence of Γ-translates of $\tilde{l}(x,c)$ for any $c\in\mathbf{R}$ converging to $\tilde{l}(x,0)$. When $c\neq kx+m$ for $k,m\in\mathbf{Z}$, this sequence corresponds in $\tilde{T}$ to a sequence of translates of the geodesic $\tilde{\lambda}(x,c)$. While the translates of $\tilde{l}(x,c)$ will come ever closer to the origin the translates of $\tilde{\lambda}(x,c)$, although following the same path through the grid $\mathscr{L}_{\#}$, must always lie outside a fixed Euclidian neighborhood of the origin. This is a consequence of Theorem 1.5. Thus the actual geodesic which these translates converge to must detour about the origin. The limit geodesic depends on whether the translates of $\tilde{\lambda}(x,c)$ approach from above or below. We denote the two limit geodesics on T by $\lambda_1(x,0)$ and $\lambda_2(x,0)$. Their symbolic dynamics are derived from $\Sigma(x,0)$ by detouring $\tilde{l}(x,0)$ about the origin. They are respectively $\Sigma^1(x,0)$ and $\Sigma^2(x,0)$.

One may argue more rigorously by considering the convergence of the functions σ_n determined by the lines $l(x,c_n)$ where $c_n\neq kx+m$ for $k,m\in\mathbf{Z}$ and $\lim_{n\to\infty}c_n=0$. The two functions $\Sigma^1(x,0)$ and $\Sigma^2(x,0)$ then arise as limits depending on whether $c_n>0$ or $c_n<0$.

The proof will be complete if we show that every minimal lamination in $\mathscr{GL}_0$ is $\lambda(x)$ or $\mathscr{L}(x)$ for some $x\in\hat{\mathbf{R}}$.

Let $\mathscr{L}$ be a minimal lamination in $\mathscr{GL}_0$. Choose a sequence of simple closed geodesics, which we can write as $\lambda(x_i)$ with $x_i \in \hat{\mathbf{Q}}$, converging to $\mathscr{L}$. Let x_i also denote a convergent subsequence. If $\lim_{i\to\infty} x_i = x$ is an irrational number then choose lines $\bar{l}(x_i, c_i)$ with c_i irrational so that $\lim_{i\to\infty} c_i = c$ is not of the form $kx+m$ for $k, m \in \mathbf{Z}$. It follows that the lifts $\tilde{\lambda}(x_i, c_i)$ realizing the same paths through the grid $\mathscr{L}_\#$ as the lines $\bar{l}(x_i, c_i)$ will converge to a geodesic $\tilde{\lambda}(x, c)$. Thus the $\lambda(x_i)$ converge to $\mathscr{L}(x)$.

If x is rational then choose irrational numbers c_i so that $\lim_{i\to\infty} c_i = c$ is irrational. As above the geodesics $\lambda(x_i)$ must converge to $\lambda(x)$. Q.E.D.

§6.3. *Proof of Theorem* 1.8. The transformations

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

pair opposite sides of the fundamental domain $F_{\mathbf{Z}}$ for the group Γ' (see Figure 6.1). Consider the transformation M_k in Γ' defined by the product $M_k = \Pi_{n=0}^{k} (B^{q(x,c)(n)} A)$ where $1 \leq x < \infty$ and $c \neq mx + l$, $m, l \in \mathbf{Z}$. M_k maps $F_{\mathbf{Z}}$ so that it intersects the $F_{\mathbf{Z}}$-segment $\tilde{\lambda}_j$, with $j = k+1+\Sigma_{n=0}^{k} q(x, c)(n)$, along a geodesic $\tilde{\lambda}$ covering the geodesic $\lambda(x, c)$. As k goes to infinity the Euclidian diameter of $M_k(F_{\mathbf{Z}})$ must approach zero. It follows, as in [3], that $\xi(x, c) = \lim_{k\to\infty} M_k(1)$ is one of the endpoints of the geodesic $\tilde{\lambda}$.

It is well known that the continued fraction expansion of a number $\xi = [a_0, a_1, a_2, \ldots]$ can be expressed in terms of the transformations

$$R = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad L = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

where

$$\lim_{n\to\infty} (R^{a_0} L^{a_1} R^{a_2} \ldots L^{a_{2n-1}} R^{a_{2n}})(1) = \xi \quad [10].$$

This is very convenient since we can write A and B as $A = LR$ and $B = RL$. Substituting R and L into the product M_k gives

$$\begin{aligned} M_k &= \prod_{n=0}^{k} (B^{q(x,c)(n)} A) \\ &= \prod_{n=0}^{k} ((RL)^{q(x,c)(n)} (LR)) \\ &= (RL)^{q(0)-1} \left(\prod_{n=1}^{k} (RL^2R^2L)(RL)^{q(n)-2} \right) (RL^2R) \end{aligned}$$

where we have written $q(x, c)(n) = q(n)$.

Since $\xi(x, c)=\lim_{k\to\infty} M_k(1)$ the above gives an explicit determination of the continued fraction expansion for $\xi(x, c)$. Using Dickensons notation [7] as in § 1, it has the form

$$1_{2q(0)-1}\, 2_2\, 1_{2q(1)-2}\, 2_2 \dots 2_2\, 1_{2q(n)-2}\, 2_2 \dots$$

When $c=0$ this approach will produce an endpoint $\xi^2(x, 0)$ of a lift of the geodesic $\lambda_2(x, 0)$ to $\mathbf{H}$. In a similar fashion one shows that a lift of the geodesic $\lambda_1(x, 0)$ has an endpoint $\xi^1(x, 0)$ of the form

$$1_{2q(0)-3}\, 2_2\, 1_{2q(1)-2}\, 2_2 \dots 2_2\, 1_{2q(n)-2}\, 2_2 \dots$$

when $2<x<\infty$ and

$$1\, 2\, 1_{2q(1)-2}\, 2_2\, 1_{2q(2)-2}\, 2_2 \dots 2_2\, 1_{2q(n)-2}\, 2_2 \dots$$

when $1<x<2$.

For $x=\infty$ it is possible to lift $\lambda(x)$ to the fixed axis of B where one of the endpoints $\xi(\infty)$ has the continued fraction expansion $[1, 1, 1, \dots]$.

We need to show that every endpoint ξ of a lift of a simple geodesic in $\bar{S}$ to $\mathbf{H}$ has the form $[b_0, b_1, \dots, b_k, a_n, a_{n+1}, \dots]$ or $[a_n, a_{n+1}, \dots]$ where $[a_0, a_1, a_2, \dots]$ is one of the numbers $\xi(x, c)$, $\xi^1(x, c)$, or $\xi^2(x, c)$ determined above with $1\leq x\leq\infty$. This will follow from elementary continued fraction theory if we show that every such ξ is of the form $g(\eta)$ where

$$g\in E=\left\{\begin{pmatrix} a & b\\ c & d\end{pmatrix} \middle| a, b, c, d\in\mathbf{Z}, ad-bc=\pm 1\right\}$$

and η is one of $\xi(x, c)$, $\xi^1(x, c)$, or $\xi^2(x, c)$ with $1\leq x\leq\infty$.

Let $\tilde{T}_{\mathbf{Z}}=\mathbf{C}\setminus\{m+n\, e^{i\pi/3} \mid m, n\in\mathbf{Z}\}$ and let $\Gamma_{\mathbf{Z}}$ be the group of translations of the form $z\mapsto z+m+n\, e^{i\pi/3}$, m, $n\in\mathbf{Z}$. $\Gamma_{\mathbf{Z}}$ acts discontinuously on $\tilde{T}_{\mathbf{Z}}$ and the quotient is conformally equivalent to the surface $T_{\mathbf{Z}}$.

The parallelogram P with vertices 0, 1, $e^{i\pi/3}$, $1+e^{i\pi/3}$ is a fundamental domain for the action of $\Gamma_{\mathbf{Z}}$ on $\tilde{T}_{\mathbf{Z}}$. We may define the symbolic dynamics of a geodesic λ with respect to the cover $\tilde{T}_{\mathbf{Z}}$ as we did with $\tilde{T}$. Here we look at the sequence of crossings a lift $\tilde{\lambda}$ realizes with translates of sides of P. The result will again describe the symbolic dynamics of λ. We can also use the lines $\tilde{l}(y, c)$ in $\tilde{T}_{\mathbf{Z}}$, as before.

$\tilde{T}$ is mapped to $\tilde{T}_{\mathbf{Z}}$ by the linear transformation

$$\begin{pmatrix} 1 & 1/2\\ 0 & \sqrt{3}/2\end{pmatrix},$$

where we treat both as subspaces of $\mathbf{R}^2$. Thus the lines $\tilde{l}(x, c)$ on $\tilde{T}$ with $1 \leqslant x \leqslant \infty$ correspond on $\tilde{T}_{\mathbf{Z}}$ to lines $\tilde{l}(y, c)$ with $\sqrt{3} \leqslant y \leqslant \sqrt{3}/3$.

An endpoint of a lift of a geodesic λ is associated with an asymptotic end of λ. The endpoints we have computed above correspond to ends of simple geodesics $\lambda(x)$ with $1 \leqslant x \leqslant \infty$. Since a lift $\tilde{\lambda}(x)$ of $\lambda(x)$ to $\mathbf{H}$ shares its endpoints with a lift $\tilde{\tilde{l}}(x, c)$ of some line $\tilde{l}(x, c)$ we may restrict our attention to lines. Furthermore, we may work with half lines or rays since these determine exactly one endpoint. The endpoints $\xi(x, c)$, $\xi^1(x, c)$, and $\xi^2(x, c)$ correspond exactly to the ends of rays $\tilde{r}(x, c)$ which are the half lines on the lines $\tilde{l}(x, c)$ with positive real part, $1 \leqslant x \leqslant \infty$ and $0 < c < 1$. When translated to $\tilde{T}_{\mathbf{Z}}$ these are the rays $\tilde{r}(y, c)$ with $\sqrt{3}/2 \leqslant y \leqslant \sqrt{3}$ and with base point c lying on the arc joining 0 to $e^{i\pi/3}$.

Let $\text{Möb}_{\mathbf{Z}}^{+}$ be the group generated by $\text{Möb}_{\mathbf{Z}}$ and the transformation $z \mapsto -\bar{z}$. When we look at the action of $\text{Möb}_{\mathbf{Z}}^{+}$ on $\mathbf{R}$ it is clear that the maps not in $\text{Möb}_{\mathbf{Z}}$ have the form $x \mapsto (ax+b)/(cx+d)$ where $a, b, c, d \in \mathbf{Z}$, $ad-bc=-1$. Hence $\text{Möb}_{\mathbf{Z}}^{+}$ restricted to $\mathbf{R}$ is the group E.

$\text{Möb}_{\mathbf{Z}}^{+}$ projects to a group $\Gamma_{\mathbf{Z}}^{+}$ of transformations acting on $\tilde{T}_{\mathbf{Z}}$ which is generated by the $(2,3,6)$ triangle group [15] and a reflection in the line through the origin with slope $\sqrt{3}/3$. It is clear that every ray $\tilde{r}(y, c)$ in $\tilde{T}_{\mathbf{Z}}$ is $\Gamma_{\mathbf{Z}}^{+}$-equivalent to one of the rays $\tilde{r}(y', c)$ with $\sqrt{3}/3 \leqslant y' \leqslant \sqrt{3}$ whose ends correspond to one of the points $\xi(x, c)$, $\xi^1(x, c)$, or $\xi^2(x, c)$ with $1 \leqslant x \leqslant \infty$. It follows that the endpoint ξ of a lift of $\tilde{r}(y, c)$ to $\mathbf{H}$ is $\text{Möb}_{\mathbf{Z}}^{+}$-equivalent to one of these points. Q.E.D.

Acknowledgements. I would like to express my thanks to Bill Abikoff, Peter Matelski, and Perry Susskind for their help and encouragement. It was Perry who first pointed out to me the connection between points of approximation and the classical Diophantine approximation. I am also indebted to George Raney for leading me to Cohn's original paper.

References

[1] BEARDON, A., *The geometry of discrete groups*. Springer-Verlag, New York 1983.
[2] BEARDON, A. & MASKIT, B., Limit points of Kleinian groups and finite sided fundamental polyhedra. *Acta Math.*, 132 (1974), 1–12.
[3] BIRMAN, J. S. & SERIES, C., An algorithm for simple curves on surfaces. To appear.
[4] — Simple curve have Hausdorff dimension one. To appear.
[5] COHN, H., Some direct limits of primitive homotopy words and of Markoffs geodesics, in *Discontinuous groups and Riemann surfaces*, Ann. of Math. Studies, 79. Princeton Univ. Press, Princeton 1974, 81–98.
[6] — Representation of Markoffs binary quadratic forms by geodesics on a perforated torus. *Acta Arith.*, 18 (1971), 125–136.
[7] DICKSON, L. E., *Studies in the theory of numbers*. Chicago, 1930.

[8] FENCHEL, W. & NIELSEN, J., *Discontinuous groups of noneuclidean motions*. Unpublished manuscript.
[9] HAAS, A., Length spectra as moduli for hyperbolic surfaces. To appear.
[10] HURWITZ, A., Über die angenaherte Darstellungen der Zahlen durch rationale Bruche. *Math. Ann.*, 44 (1894), 417–436.
[11] JORGENSEN, T., Simple geodesics on Riemann surfaces. *Proc. Amer. Math. Soc.*, 86 (1982), 120–122.
[12] LEHNER, J. & SHEINGORN, M., Simple closed geodesics on $H^+/\Gamma(3)$ arise from the Markov spectrum. To appear.
[13] LEVITT, G., Foliations and laminations of hyperbolic surfaces. *Topology*, 22 (1983), 119–135.
[14] MARKOFF, A. A., Sur les formes binaires indefinies, I. *Math. Ann.*, 15 (1879), 281–309; II, 17 (1880), 379–400.
[15] MASKIT, B., Manuscript on Kleinian groups. To be published.
[16] NIELSEN, J., Die Isomorphismen der allgemeinen unendlichen Gruppe mit zwei Erzeugenden. *Math. Ann.*, 78 (1918), 385–397.
[17] RANKIN, R. A., Diophantine approximation and horocyclic groups. *Canad. J. Math.*, 9 (1957), 155–182.
[18] SCHMIDT, A. L., Minimum of quadratic forms with respect to Fuchsian groups, I. *J. Reine Angew. Math.*, 286/287 (1976), 341–368.
[19] SERIES, C., The infinite word problem and limit sets in Fuchsian groups. *Ergodic Theory Dynamical Systems*, 1 (1981), 337–360.
[20] SERIES, C., Personal communication.
[21] SULLIVAN, D., Disjoint spheres, approximation by imaginary quadratic numbers, and the logarithm law for geodesics. *Acta Math.*, 149 (1982), 215–239.
[22] THURSTON, W., *The geometry and topology of 3-Manifolds*. Lecture notes, Princeton 1978.
[23] GORŠKOV, D. S., Lobačevskii geometry in connection with some problems in arithmetic. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.*, 67 (1977), 39–85.

Received March 23, 1984
Received in revised form March 9, 1985