ISRAEL JOURNAL OF MATHEMATICS, Vol. 52, No. 3, 1985

ASYMPTOTIC DISTRIBUTION OF CLOSED GEODESICS

BY
MARK POLLICOTT
Mathematics Institute, University of Warwick, Coventry CV4 7AL, England

ABSTRACT
In this paper the distribution of closed geodesics on surfaces of constant negative curvature are studied from a dynamical viewpoint. Asymptotic estimates are derived independently of the work of Selberg or Margulis, or the work of Bowen on Axiom A flows.

§0. Introduction

In this paper we are concerned with geodesic flows on the unit tangent bundle of compact surfaces of constant negative curvature [2], [12]. One way to study the lengths of closed geodesics is to use the work of Selberg [15], [30]. In this paper we develop a completely different approach. We apply results for suspension flows with the aid of the symbolic dynamics for geodesic flows due to C. Series [26], [27].

In recent work Parry and the author have derived asymptotic estimates for the number of closed orbits of an Axiom A flow [24]. (In the Axiom A case the proof is heavily dependent on the remarkable work of Bowen [6].) Axiom A flows subsume the case of geodesic flows for surfaces of constant negative curvature. There is a definite loss in generality in studying only geodesic flows. However, in this special case we are able to derive additional results of a more geometric nature. But more importantly we are able to avoid the extremely complicated machinery of Markov partitions needed for Axiom A flows.

In Section 1 we recall some known results for suspension flows. As an application we give a new result relating the equilibrium state for a Hölder continuous function (on a subshift of finite type) to the distribution of periodic points. Section 2 contains a short exposition of symbolic dynamics for geodesic

Received March 13, 1984 and in revised form October 30, 1984

flows. In Section 3 we explicitly relate the zeta functions for the geodesic flow and the suspension flow associated with it. In Section 4 we recover significant results of Margulis and Bowen on the distribution of closed geodesics. In the final section we give two new results relating the distribution of closed geodesics to the topology of the surface.

I would like to thank Keith Burns and Caroline Series for helpful comments. I would also like to thank the S.E.R.C. for their financial support.

I am particularly grateful to William Parry for his encouragement and help throughout the course of this work.

§1. Closed orbits for suspended flows

Let A be an irreducible $k \times k$ zero-one matrix and define a *shift of finite type* (Σ_A, σ) by

$$\Sigma_A = \left\{ x \in \prod_{-\infty}^{\infty} \{1,\dots,k\} \,\middle|\, A(x_n, x_{n+1}) = 1 \text{ for all } n \in \mathbf{Z} \right\}$$

and $\sigma : \Sigma_A \to \Sigma_A$ by $(\sigma x)_n = x_{n+1}$.

Given $0 < \theta < 1$ we define a metric on Σ_A by $d(x,y) = \theta^n$, where $n \geqq 0$ is the largest integer such that $x_i = y_i$, $|i| \leqq n-1$. With this metric Σ_A is compact and zero-dimensional.

Assume that $f : \Sigma_A \to \mathbf{R}^+$ is Lipschitz, i.e. there exists $C > 0$ s.t.

$$|f(x) - f(y)| \leqq C d(x,y) \qquad \text{for all } x, y \in \Sigma_A.$$

The *suspension space* is defined to be the compact metrizable space

$$\Sigma_A^f = \{(x,r) \mid x \in \Sigma_A, 0 \leqq r \leqq f(x)\}$$

where $(x, f(x))$ and $(\sigma x, 0)$ are identified.

The *suspension flow* σ^f is given locally by $\sigma_t^f(x,r) = (x, r+t)$, with appropriate identifications. If we assume that

$$\sup\left\{ h(m) - \int f dm \,\middle|\, m \ \sigma\text{-invariant} \right\} = 0$$

then the topological entropy $h(\sigma^f)$ is unity [24]. (Here $h(m)$ is the entropy of σ with respect to m.)

Furthermore the measure of maximal entropy for σ^f is $\mu \times l / \int f d\mu$, where μ is the unique measure attaining the above supremum and l is linear Lebesgue measure.

The suspended flow is called (topologically) *weak-mixing* if $F\sigma_t^f = e^{iat}F$ has no non-trivial solution ($F \in C(\Sigma_A)$). A closed σ-orbit $\{x, \sigma x, \ldots, \sigma^{m-1}x\}$ (where $\sigma^m x = x$) corresponds to a closed σ^f-orbit of length

$$f^m(x) = f(x) + \cdots + f(\sigma^{m-1}x).$$

The following result was proved in [23].

PROPOSITION 1. *Let $k : \Sigma_A \to \mathbf{R}^+$ be Lipschitz and σ^f a weak-mixing suspension flow, then*

$$\sum_{m=1}^{\infty} \frac{1}{m} \sum_{\sigma^m x = x} k^m(x) e^{-sf^m(x)} = \frac{\int k d\mu / \int f d\mu}{s-1} + \phi(s) \tag{1.1}$$

where $\phi(s)$ is analytic on a neighbourhood of $\{s \mid \mathcal{R}(s) \geqq 1\}$.

This proposition has an immediate interpretation in terms of closed orbit distributions. We write $A(t) \sim B(t)$ if $A(t)/B(t) \to 1$ as $t \to \infty$. We let $O(A(t))$ denote a term whose ratio to $A(t)$ is bounded above.

PROPOSITION 2. *Let $k : \Sigma \to \mathbf{R}^+$ be Lipschitz, then*

(i) $\Sigma_{e^{f^m(x)} \leqq t} 1 \sim t/\log t$

(ii) $\Sigma_{e^{f^m(x)} \leqq t} k^m(x) \sim t \int k d\mu / \int f d\mu$

where σ^f is weak-mixing (summations are over all periodic orbits $\{x, \sigma x, \ldots, \sigma^{m-1}x\}$).

PROOF. Denote

$$F_k(t) = \Sigma_{e^{f^m(x)} \leqq t} k^m(x).$$

We can rewrite (1.1) using a Riemann–Stieltjes integral

$$\int_1^{\infty} t^{-s} dF_k(t) = \frac{\int k d\mu / \int f d\mu}{s-1} + \phi(s).$$

The Ikehara–Wiener Tauberian theorem [31] then gives that $F_k(t) \sim t \int k d\mu / \int f d\mu$. This completes the proof of (ii).

For part (i) take

$$G(t) = \Sigma_{e^{f^m(x)} \leqq t} 1.$$

Then we can write

$$G(t) = \int_2^t \frac{1}{\log r} dF_f(r) + O(1).$$

We can rewrite this as follows:

$$G(t) = \left[\frac{F_f(r)}{\log r}\right]_2^t + \int_2^t \frac{F_f(r)}{r(\log r)^2}dr$$

$$\sim \frac{t}{\log t} + \int_2^t \frac{r}{r(\log r)^2}dr$$

$$\sim \frac{t}{\log t} - \left[\frac{r}{\log r}\right]_2^t + \int_2^t \frac{1}{\log t}dt$$

$$\sim \mathrm{li}(t)$$

$$\sim \frac{t}{\log t}$$

(since $\mathrm{li}(t) = \int_2^t (1/\log t)dt \sim (t/\log t)$). This completes the proof.

REMARKS. (i) By a similar manipulation of Riemann–Stieltjes integrals other asymptotic results can be deduced, e.g.

$$\Sigma_{e^{f^m(x)} \leqq t}\, e^{-f^m(x)} = \int_2^t \frac{1}{r}dG(r) + O(1)$$

$$= \int_2^t \frac{1}{r\log r}dr + O(1)$$

$$= \log\log t + O(1).$$

Then following the arguments of [14], §22.18 we can prove

$$\Sigma_{e^{f^m(x)+f^n(y)} \leqq t}\, 1 \sim \frac{t}{\log t}\log\log t,$$

where the summation is over pairs of closed orbits $\{x, \sigma x, \ldots, \sigma^{m-1}x\}$, $\{y, \sigma y, \ldots, \sigma^{n-1}y\}$.

(ii) The error terms for the asymptotic estimates depend on the extension of $\phi(s)$ to $\mathcal{R}(s) \leqq 1$. For geodesic flows (which we shall see give rise to suspension flows) the error is $O(t^\alpha)$, for some $0 < \alpha < 1$, in Proposition 2 ([15], p. 64). For certain locally constant functions, however, this can never be the case (cf. [18] Ch. 5 and [25] §5).

(iii) William Parry has shown the author that (1.1) may be generalised to:

$$\sum_{m=1} \frac{1}{m} \sum_{\sigma^m x = x} \frac{(g_1^m \cdots g_k^m)(x)}{[f^m(x)]^{k-1}} e^{-sf^m(x)} = \frac{\int g_1 d\mu \cdots \int g_k d\mu}{[\int f d\mu]^k} \cdot \frac{1}{s-1} + \phi(s)$$

where $g_1, \ldots, g_k$ are Lipschitz and $\phi(s)$ is analytic on a neighbourhood of $\{s \mid \mathcal{R}(s) \geqq 1\}$.

APPLICATION (of Proposition 2). Let μ_0 be the measure of maximal entropy for $\sigma:\Sigma_A\to\Sigma_A$. It is well known that if $Q_0(t)=\{x \mid \sigma^n x = x,\ n\leqq t\}$ then $(1/\mathrm{Card}\, Q_0(t))\Sigma_{x\in Q_0(t)}\delta_x$ converges to μ_0 in the weak* topology (here δ_x is the measure consisting of a single atom at x; [1] p. 302). We will now derive a natural generalisation of this result to equilibrium states.

Let $h:\Sigma_A\to\mathbf{R}$ be Lipschitz and denote its pressure by $P(h)$. By replacing h by $h-P(h)$ we may assume that $P(h)=0$. Furthermore, by adding a coboundary we may assume that $f=-h>0$. (This requires the Ruelle operator theorem; cf. [24].) If μ is the unique equilibrium state for h then this is unaltered by the above changes.

Assume σ^f is weak-mixing. Define

$$Q(t)=\{x \mid \sigma^n x = x,\ nP(h)-h^n(x)\leqq t\},$$

then by Proposition 2, $(1/\mathrm{Card}\, Q(t))\Sigma_{x\in Q(t)}\delta_x$ converges to μ in the weak* topology. (For the purposes of this proof we are assuming that $P(h)=0$.)

If σ^f is not weak-mixing then it can be represented (after recoding) by a constant suspension. Therefore the weak* approximation of maximal measures again shows that $(1/\mathrm{Card}\, Q(t))\Sigma_{x\in Q(t)}\delta_x$ converges to μ.

In either event we have proved the following result.

COROLLARY 2.1. *Let μ be the unique equilibrium state for $h:\Sigma_A\to R$ and let $Q(t)$ be as defined above, then $(1/\mathrm{Card}\, Q(t))\Sigma_{x\in Q(t)}\delta_x$ converges to μ in the weak* topology as $t\to+\infty$.*

§2. Symbolic dynamics for geodesic flows

Let M be a compact Riemann surface of curvature $\kappa=-1$ (and genus $\mathfrak{g}\geqq 2$). Topologically M is a $\mathfrak{g}$-holed torus. Define the geodesic flow $\phi_t: T_1M\to T_1M$ on the unit tangent bundle as follows: Given $(x_0,v_0)\in T_1M$ let $\gamma:\mathbf{R}\to M$ be the unique (unit speed) geodesic passing through x_0, in direction v_0, at time $t=0$, i.e. $\gamma(0)=x_0$ and $\dot\gamma(0)=v_0$. Then define $\phi_t(x_0,v_0)=(\gamma(t),\dot\gamma(t))$. Thus closed ϕ-orbits in T_1M correspond to closed geodesics in M.

Let D be the Universal covering space for M and $\pi:D\to M$ the projection. Let Γ be the group of covering transformations (i.e. $g\in\Gamma$ is an isometry $g:D\to D$ s.t. $\pi g=\pi$). Since $\mathfrak{g}\geqq 2$, the Universal cover D is the Poincare disc $\{z \mid |z|<1\}$ with metric

$$dx^2+dy^2=\frac{dr^2}{(1-r^2)^2}$$

(where dx, dy are increments in the hyperbolic metric, and dr is an increment in the Euclidean metric) [20], [21].

With respect to this metric the geodesics in D are circular arcs meeting S^1 orthogonally [3].

Since Γ is a finitely generated, discontinuous group of isometries on D it forms a Fuchsian group. Elements $g \in \Gamma$ are necessarily linear fractional transformations of the form

$$g(z) = \frac{az + b}{\bar{b}z + \bar{a}},$$

where $|a|^2 - |b|^2 = 1$ [13]. Let Γ_0 be a finite set of generators for Γ, together with their inverses.

To find a canonical representation for M in D look at the loci $\{z \mid |g'(z)| = 1\} = C(g)$, where $g \in \Gamma_0$. These are geodesic arcs in D called isometric circles. If $g \in \Gamma_0$ then $gC(g) = C(g^{-1})$ [13]. Furthermore g takes the "interior" of $C(g)$ to the "exterior" of $C(g^{-1})$ (Fig. 1).

Since Γ is the group of covering transformations we may identify M and D/Γ. The region R exterior to all of the isometric circles $C(g)$, $g \in \Gamma_0$, is called the *Fundamental region.* Since R is a "maximal" set for which two points are not identified it represents M in the covering space D. (Label the elements of Γ_0 by $g_1, g_2, \dots$ so that the isometric circles $C(g_1), C(g_2), \dots$ encircle R in an anticlockwise direction.) (See Fig. 2; the isometric circles are arranged so that R is bounded away from S^1.)

Other representations of M are given by gR, $g \in \Gamma$. Together these "tile" the disc D by representations all of the same size (in the hyperbolic metric).

By following images of O (under Γ) around any vertex of R we get the unique defining relation for Γ ([13] §27); see Fig. 3.

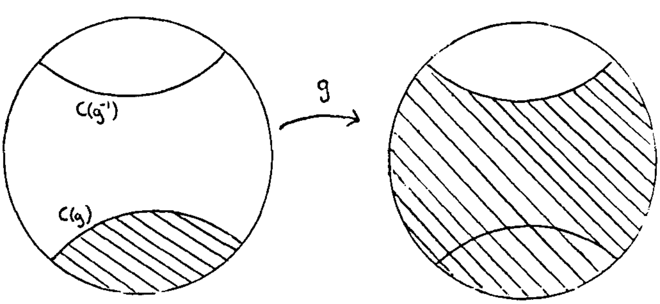

Fig. 1.

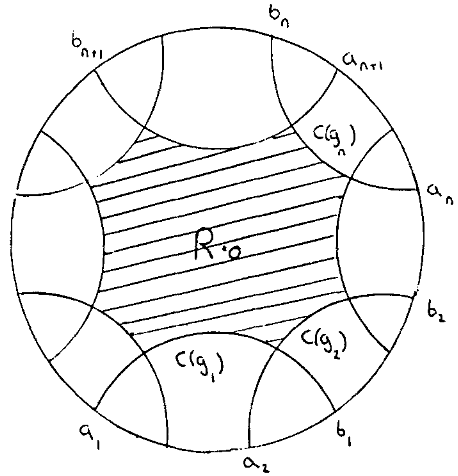

Fig. 2.

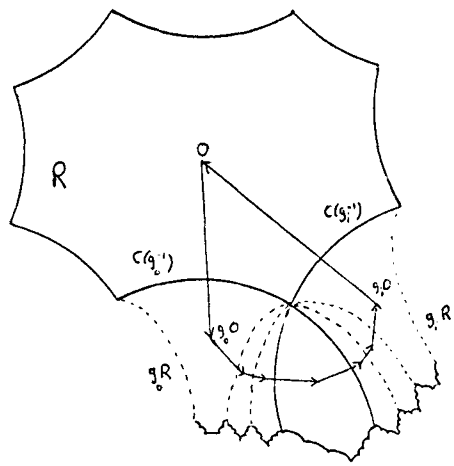

Fig. 3.

A simplifying assumption is:

(*) For each $g \in \Gamma_0$, $C(g)$ is covered by images of ∂R under Γ [8], [26].

Condition (*) holds in the standard "symmetrical" case where R is a regular figure.

Consider the piecewise C^2 map $T: S^1 \to S^1$ given by $T(z) = g_i(z)$, $z \in [a_i, a_{i+1})$. (Here $i+1$ is given modulo 4g.) For large N, $|(T^N)'| > 1$ and given $x \in S^1$ we may choose a sequence $(g_{i(n)}^{-1})_{n=0}^{\infty}$ so that $T^n(x) \in [a_{i(n)}, a_{i(n)+1}]$. We shall call this sequence a T-expansion for x.

Similarly for the map $\bar{T}(z)=g_i(z)$, $z\in(b_i,b_{i+1}]$. Here $y\in S^1$ gives rise to a sequence $(g_{j(n)}^{-1})$ where $\bar{T}^n(y)\in[b_{j(n)},b_{j(n)+1}]$. We call this sequence a $\bar{T}$-expansion for x.

THEOREM 1 (Series [26]). *There exists a subshift (not of finite type) $\Sigma\subseteq\Pi_{-\infty}^{+\infty}\Gamma_0$, a Hölder continuous function $h:\Sigma\to\mathbf{R}^+$, and a continuous surjection $\pi:\Sigma^h\to T_1M$ such that $\pi\sigma_t^h=\phi_t\pi$ and π is bounded-to-one.*

A sequence $w=(g_{r(n)})_{n=-\infty}^{\infty}$ (where $w\in\Sigma$) gives rise to expansions for two points $x,y\in S^1$. (Let $(g_{r(n)})_{n=1}^{\infty}$ represent a T-expansion for x, and $(g_{r(-n)}^{-1})_{n=0}^{\infty}$ be a $\bar{T}$-expansion for y.) These two points on S^1 uniquely determine a geodesic γ on D. Flowing from $(w,0)$ to $(w,h(w))$ corresponds to the unit tangent vector moving along $\gamma\cap\bar{R}$. (Complications arise if $\gamma\cap\bar{R}=\varnothing$, when R is replaced by another representation gR, $g\in\Gamma$.)

NOTE. Series shows that the condition (*) is not essential and a surface which satisfies this condition can be related to one which does not by a quasi-conformal mapping ([26] §6). In particular, if Σ is constructed for the standard "symmetrical" case of the same genus, then h,π can be chosen so that Theorem 1 is still true even when the surface does not satisfy condition (*).

The measure of maximal entropy for ϕ is the Riemann measure ([5] p. 420). Furthermore the geodesic flow ϕ is topologically weak-mixing [11].

§3. Closed orbits and zeta functions

To use the results of §1 we need to replace Σ by a subshift of finite type Σ_A.

Bowen and Series constructed a Markov partition for $T:S^1\to S^1$ as follows: For each vertex v_i of R let $N(v_i)$ be those geodesic arcs passing through v_i which are images of $C(g)$ $(g\in\Gamma_0)$ under Γ [8] (see Fig. 4).

Let W be the points where arcs in $\bigcup_i N(v_i)$ meet S^1. If we still assume (*) then $\{a_i\},\{b_i\}\subseteq W$. Furthermore $TW\subseteq W$ (here we could take $f(a_i)$ to be either $g_i(a_i)$ or $g_{i-1}(a_i)$ and this would still be true) [8].

The points W therefore divide S^1 up into a Markov partition. Let Σ_A^+ be the one-sided shift of finite type given by this Markov partition. Since this new partition refines the old one there is an obvious map $i:\Sigma_A^+\to\Sigma^+$ (where Σ^+ is the one-sided subshift corresponding to Σ). This extends to $i:\Sigma_A\to\Sigma$. We want to compare closed orbits in $\Sigma_A^{h\circ i}$ and T_1M. There will be a one-one correspondence unless a T_1M closed orbit has an Σ_A expansion which is not unique. By periodicity it suffices to compare closed orbits for $\sigma:\Sigma_A^+\to\Sigma_A^+$ and $T:S^1\to S^1$. In this case the only difficulty is when $x\in W$.

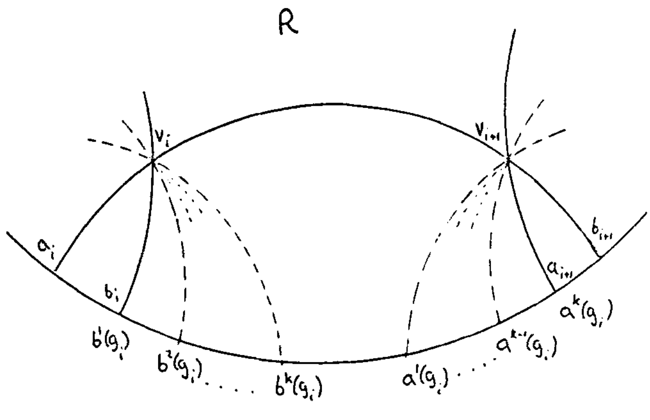

Fig. 4.

Let

$$I_1^i = \{x \in W \mid x \in N(v_i) \cap (a_i, a_{i+1}]\},$$
$$I_2^i = \{x \in W \mid x \in N(v_{i+1}) \cap [a_i, a_{i+1}]\}$$

and

$$I_1 = \bigcup_i I_1^i, \qquad I_2 = \bigcup_i I_2^i.$$

Next let $I_1^i = \{b^1(g_i), \ldots, b^k(g_i)\}$ and $I_2^i = \{a^1(g_i), \ldots, a^k(g_i)\}$, where, in particular, $b^1(g_i) = b_i$ and $a^k(g_i) = a_{i+1}$. The elements of I_1^i, I_2^i are listed in the order in which they are labelled around S^1. (Here $k = 2g - 1$.) Because of these definitions $g_i I_1^i \subseteq I_1$ and $g_i I_2^i \subseteq I_2$. Furthermore

$$\begin{cases} Ta^r(g_i^{-1}) = a^{r-1}(g_j) & \text{(for some } j = j(r,i)) \\ Tb^r(g_i^{-1}) = b^{r+1}(g_{j'}) & \text{(for some } j' = j'(r,i)) \end{cases}$$

(in fact, $j(r,i) = i - 1$ $(r \neq 1)$, $j(1,i) = i - 2$ and $j'(r,i) = i + 1$ $(r \neq k)$, $j'(k,i) = i + 2)$, where j, j' and $r, r-1$ and $r+1$ are given modulo $4g$ and modulo k, respectively.

NOTE. The classical case is where the isometric circles have the clockwise ordering

$$g_{4g} \cdots g_4 g_3 g_2 g_1 = p_1 q_1 p_1^{-1} q_1^{-1} \cdots p_g q_g p_g^{-1} q_g^{-1}$$

(cf. [21]), where we have simply relabelled the generation g_i by $p_j^{\pm 1}$, $q_j^{\pm 1}$.

In this case g_{i+1} follows g_i^{-1} in the cyclic sequence

$$q_g^{-1}p_gq_gp_g^{-1}\cdots q_1^{-1}p_1q_1p_1^{-1}.$$

(These two sequences are precisely the surface symbol and the defining relation [21].)

If g is even then we have that I_1 contains 2g T-orbits of period $2(2g-1)$. If g is odd then I_1 contains 4 orbits of period $g(2g-1)$.

Points in I_1 lie between two intervals in the Markov partition. A consistent choice of intervals to the left or to the right for each T-orbit (in I_1) gives rise to two distinct T-expansions and hence two distinct closed σ^h-orbits. The set I_2, however, does not give rise to any additional orbits in Σ_A. This is because the choice of intervals to the right breaks down at a_i. Thus the primitive closed orbits for ϕ and $\sigma^{h\circ i}$ differ by only a finite number l of orbits of length C (corresponding to σ-orbits of period p), say. (Here p and l depend only on g, the genus of the surface. The actual value of these constants will prove to be unimportant.) Providing we have chosen h in accordance with ([26] §5) for a surface not satisfying (*), those ϕ and $\sigma^{h\circ i}$ orbits in one-one correspondence will be of the same period.

Comparing zeta functions we see that

$$\zeta_\phi(s)=\prod_\tau(1-e^{-s\lambda(\tau)})^{-1}$$

(here the product is over closed orbits τ of length $\lambda(\tau)$)

$$\begin{aligned}&=\exp\left(\sum_{m=1}^{\infty}\frac{1}{m}\sum_{\sigma^m x=x}e^{-sh\circ i^m(x)}-l\sum_{n=1}^{\infty}\frac{e^{-snC}}{n}\right)\\&=\zeta_{\sigma^{h\circ i}}(s)\exp l\log(1-e^{-sC})\\&=\zeta_{\sigma^{h\circ i}}(s)(1-e^{-sC})^l.\end{aligned}$$

REMARKS. (i) Let $G(z)$ be the generating function for homotopy class word lengths, then

$$G(z)=\sum_{n=1}^{\infty}z^n\operatorname{Card}\{\text{Homotopy classes length } n\}.$$

Cannon has shown that $G(z)$ is rational [10]. Let us consider the generating function for free homotopy classes $G_0(z)$. Every free homotopy class corresponds to exactly one closed geodesic ([4] VII.6). Because of the ambiguity in representing elements we shall take the length of a free homotopy class to be the word length of the corresponding closed geodesic

$$\begin{aligned} G_0(z) &= \sum_{n=1}^{\infty} z^n \operatorname{Card}\{\text{Free homotopy classes length } n\} \\ &= \sum_{n=1}^{\infty} \frac{z^n}{n} \operatorname{Card}\{x \mid \sigma^n x = x\} - l \sum_{n=1}^{\infty} \frac{z^{np}}{n} \\ &= \log\left[\frac{(1-z^p)^l}{\det(I - zA)}\right]. \end{aligned}$$

In particular, $\exp G_0(z)$ is rational.

(ii) Let $\alpha > 0$ be the growth rate of cyclically reduced words, then α is also the growth rate of word lengths of closed geodesics. Then e^α is a pole for $G_0(z)$ and so a zero of the polynomial $\det(I - zA)$. Thus e^α is an algebraic number.

Notice that since ϕ and $\sigma^{h \circ i}$ have closed orbits of the same lengths, $\sigma^{h \circ i}$ is weak-mixing [7].

§4. Asymptotic results for closed geodesics

In [16], Hejhal gives an asymptotic formula for the number of closed geodesics on a surface of constant negative curvature. We shall now give a proof with a more dynamic flavour.

Let $v(t)$ be the number of closed geodesics of length at most t. Then $v(t) = v_\phi(t)$, the number of closed ϕ-orbits of period at most t.

THEOREM 2 (Margulis, cf. Hejhal [15], [16]). *Let M be a compact surface of curvature $\kappa = -1$, then $v(t) \sim e^t/t$.*

PROOF. We saw in the previous section that $\zeta_\phi(s) = \zeta_{\sigma^{h \circ i}}(s)(1 - e^{-sC})^l$ and in particular there are only l additional primitive closed orbits for $\sigma^{h \circ i}$ compared with ϕ (each of length C). Thus $v_\phi(t)$ and $v_{\sigma^{h \circ i}}(t)$ differ only by iterates of these orbits, i.e.

$$v_{\sigma^{h \circ i}}(t) = v_\phi(t) + l[t/C].$$

By Proposition 2, $v_{\sigma^{h \circ i}}(t) \sim e^t/t$ thus $v_\phi(t) \sim e^t/t$. This completes the proof.

We can also approach the problem of distribution of closed orbits on T_1M. Let B be a subset of T_1M whose boundary has measure zero. Let $v_B(t)$ be the *total* sojourn time in B of closed orbits whose periods are at most t.

THEOREM 3 (Bowen [5], cf. Parry [23]). *Let M be a compact surface of curvature $\kappa = -1$, and let $B \subseteq T_1M$ have $m(\partial B) = 0$, then*

$$v_B(t) \sim \frac{m(B)}{m(M)} e^t/t.$$

Proof. We already know this result is true for suspension flows and that the scalar $m(B)/m(M)$ is appropriate. Again the only difference is due to a finite number of orbits. This contributes a difference of order t. Therefore the result follows.

Remark. The above theorem shows that closed orbits are uniformly distributed in T_1M (with respect to m). This means that closed geodesics are uniformly distributed on M (with respect to Riemann measure) and in direction (with respect to Lebesgue measure).

Remark. The method we have used only applies to compact surfaces. A special case of a non-compact surface is the modular surface (corresponding to $\mathrm{SL}(2, Z)$). In this instance closed geodesics have the same lengths as closed orbits for a suspension over the continued fraction transformation, where we can take $h(x) = -2\log x$ [29]. Then work of Mayer on the corresponding zeta function would allow us to extend Theorem 2 to this special case [22].

§5. Further distribution results

By following the formulation due to C. Series we are better able to describe the distribution of closed orbits with respect to the topological structure. Under the assumption (*) the sides of R form closed loops from arcs of closed geodesics. Topologically these loops ring the holes in the surface or pass through such holes. We can interpret a geodesic as passing through a hole if it cuts the first type of loop. Similarly we may interpret a geodesic as circling a hole if it traverses the second type of loop (see Fig. 5).

For a closed geodesic τ let $w(\tau)$ be the total number of such crossings by τ. Assume that the corresponding $\sigma^{h\circ i}$-orbit lies over a σ-periodic point $\sigma^n x =$

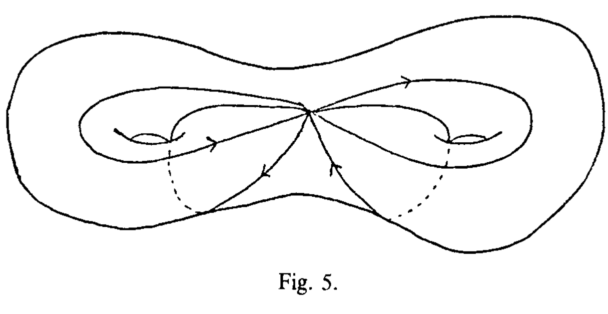

Fig. 5.

$x \in \Sigma_A$. Then the period of x is the number of crossing of R, say, i.e. $w(\tau) = n$. For $g \in \Gamma_0$ let $w_g(\tau)$ be the number of crossings of the loop corresponding to $C(g) \cap \bar{R}$.

THEOREM 4. *Let M be a compact surface with curvature $\kappa = -1$ and satisfying* (*), *then*:

(i) *There exists $C > 0$ such that $\Sigma_{\lambda(\tau) \leq t} w(\tau) \sim Ce^t$.*

(ii) *For $g \in \Gamma_0$ there exists $C_g > 0$ such that*

$$\sum_{\lambda(\tau) \leq t} w_g(\tau) \sim C_g \cdot e^t.$$

PROOF. (i) By Proposition 2(ii) we know that

$$\sum_{h \circ i^n(x) \leq t} n \sim \frac{e^t}{\int f d\mu}$$

(where we take $k : \Sigma_A \to \mathbf{R}$ to be the constant function 1). Equating $h \circ i^n(x)$ with $\lambda(\tau)$ and n with $w(\tau)$ we have the result (the additional primitive closed orbits making no difference to the asymptotic formula).

(ii) Let $B_g \subseteq \Sigma$ be those sequences representing geodesics crossing $C(g) \cap \bar{R}$, say. Let $C_g = \mu(B_g)/\int h d\mu$. The result follows from Proposition 2(ii) by taking $k = \chi_{B_g}$ (after extending Proposition 2 to $k = \chi_{B_g}$ by approximation).

REMARK. If $C(g) \cap \bar{R}$ has length $l(g)$ then $C(g) = l(g)/\pi \cdot \mathrm{vol}\, M$. If ∂R has length $l(\partial R)$ then $C = l(\partial R)/\pi \cdot \mathrm{vol}\, M$.

The above theorem relates lengths of closed geodesics to their intersections with fixed geodesics $C(g)$. We want to adapt these ideas to accommodate self-intersections.

Let M be a manifold of infinite volume where generators in Γ_0 satisfy $C(g_i) \cap C(g_j) = \emptyset$ with $g_i, g_j \in \Gamma_0$ $(g_i \neq g_j)$ [17], [28] (see Fig. 6). A geodesic γ in D is labelled by a bi-infinite sequence from Γ_0. Let γ cut $C(g_0)$, $g_0^{-1}C(g_1)$, $g_0^{-1}g_1^{-1}C(g_2)$, $g_0^{-1}g_1^{-1}g_2^{-1}C(g_3), \ldots$, where we start in R. Furthermore, going in the opposite direction assume γ cuts $C(g_{-1}^{-1})$, $g_{-1}C(g_{-2}^{-1})$, $g_{-2}g_{-1}C(g_{-3}^{-1}), \ldots$. We represent γ by the sequence (g_n^{-1}) [28].

If the geodesic γ lies in the non-wandering set Ω (which includes all closed geodesics) then the sequence does not terminate. The only admissibility condition is that g is not followed by g^{-1}. Admissible sequences form a shift of finite-type Σ_A. For a geodesic γ with a sequence $x \in \Sigma_A$ let $r(x)$ be the hyperbolic length of $\gamma \cap \bar{R}$. Then $r : \Sigma_A \to \mathbf{R}^+$ is Hölder continuous. There is a topological conjugacy $\pi : \Sigma_A^r \to T_1\Omega$ s.t. $\pi\sigma_t^r = \phi_t\pi$.

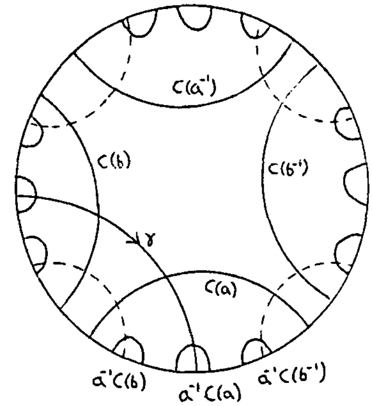

Fig. 6.

In order that a geodesic on M intersects itself it must have distinct lifts to D which cross in R. The number of self-intersections is exactly the number of such intersections in D. The most obvious case is if liftings γ and γ' enter and leave R by isometric circles which alternate around S^1, i.e. if γ, γ' have expansions x, y then $x_0 \neq y_0$ and $x_{-1} \neq y_{-1}$ and the pairs $C(x_0^{-1})$, $C(x_{-1})$ and $C(y_0^{-1})$, $C(y_{-1})$ separate each other around S^1. We may refine this: Let $w = [w_{-n}, \dots, w_{n-1}]$ and let $P(w)$ be the set of $2n$-cylinders which differ from w only at both ends (in $-n$ and $n-1$ co-ordinates), i.e. for $w' = [w'_{-n}, \dots, w'_{n-1}] \in P(w)$; $w'_{-n} \neq w_{-n}$, $w'_{n-1} \neq w_{n-1}$, $w_i = w'_i$ $(-n < i < n-1)$. Furthermore, impose the condition on $P(w)$ that the circular arcs $w_0 w_1 \cdots C(w_{n-1}^{-1})$, $w_{-1}^{-1} w_{-1}^{-1} w_{-2}^{-1} \cdots C(w_{-n})$ separate $w'_0 w'_1 \cdots C(w'^{-1}_{n-1})$ and $w'^{-1}_1 w'^{-1}_2 \cdots C(w'_{-n})$ (cf. [8]), and there exist geodesics γ, γ' in D which cross within R and whose corresponding sequences are contained in w, w' respectively. If x defines a geodesic, then a self-intersection occurs if $x \in w$ and $\sigma^k x \in w' \in P(w)$. We can count the number of *such* self-intersections (given $\sigma^m x = x$) as

$$S_n(x) = \frac{1}{2} \sum_{w} \sum_{w' \in P(w)} [\chi_w]^m(x) [\chi_{w'}]^m(x),$$

where χ_w is a characteristic function and n is fixed.

The number of self-intersections per unit length is $S_n(x)/r^m(x)$. Consider

$$\eta(s) = \sum_{\substack{w \\ w' \in P(w)}} \left\{ \sum_{m=1}^{\infty} \frac{1}{m} \sum_{\sigma^m x = x} \frac{[\chi_w]^m(x) [\chi_{w'}]^m(x)}{2 \cdot r^m(x)} e^{-s r^m(x)} \right\}$$

$$= \frac{l_n}{s-1} + \psi(s)$$

where

$$l_n = \frac{1}{2\int r d\mu} \sum_{\substack{w \\ w' \in P(w)}} \mu(w) \cdot \mu(w'),$$

and μ is the equilibrium state for $-h$ (cf. §1). Thus an nth approximation to counting self-intersections of closed orbits is

$$\sum_{\lambda(\tau) \leqq t} \frac{S_n(\tau)}{\lambda(\tau)} \sim l_n e^t.$$

We now consider the limit of these approximations. Let $S(\tau)$ be the number of self-intersections of τ. Define $B \subseteq (S^1 \times S^1) \times (S^1 \times S^1)$ by

$$B = \{(x_1, x_2; y_1, y_2) : x_i < y_j < x_k < y_l \text{ lie in order around } S^1; \{i, k\} = \{j, l\} = \{1, 2\} \text{ and the corresponding geodesics cross in } R\}.$$

Recall that $\pi : \Sigma_A \to S^1 \times S^1$ and thus $\pi\mu$ is a measure on $S^1 \times S^1$.

THEOREM 5.

$$\sum_{\lambda(\tau) \leqq t} \frac{S(\tau)}{\lambda(\tau)} \sim (\pi\mu \times \pi\mu)(B) \cdot e^t/2.$$

This follows by approximation. Thus the average number of self-intersections per unit length is $(\pi\mu \times \pi\mu)(B)/2$.

REFERENCES

1. V. M. Alexeev and M. V. Jacobson, *Symbolic dynamics and hyperbolic dynamic systems,* Phys. Rep. **75** (1981), 287–325.
2. V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics,* Benjamin, New York, 1968.
3. A. F. Beardon, *Geometry of discrete groups,* in *Discrete Groups and Automorphic Functions* (W. J. Harvey, ed.), Academic Press, London, 1977.
4. M. Berger, *Lectures on Geodesics in Riemannian Geometry,* Tate Institute, Bombay, 1965.
5. R. Bowen, *Equidistribution of closed geodesics,* Am. J. Math. **94** (1972), 413–423.
6. R. Bowen, *Symbolic dynamics for hyperbolic flows,* Am. J. Math. **95** (1973), 429–459.
7. R. Bowen, *Mixing Anosov flows,* Topology **15** (1976), 77–79.
8. R. Bowen and C. Series, *Markov maps associated with Fuchsian groups,* Publ. Math. I.H.E.S. **50** (1979), 153–170.
9. J. S. Birman and C. Series, *An algorithm for simple curves on surfaces,* preprint.
10. J. W. Cannon, *The growth of the closed surface groups and the compact hyperbolic Coxeter groups,* preprint.
11. P. Eberlien, *Geodesic flows on negatively curved manifolds II,* Trans. Am. Math. Soc. **178** (1973), 57–82.

12. S. V. Fomin and I. M. Gelfand, *Geodesic flows on manifolds of constant negative curvature*, Transl. Am. Math. Soc. (2) **1** (1955), 49–65.

13. L. R. Ford, *Automorphic Functions*, McGraw-Hill, New York, 1929.

14. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, Oxford, 1938.

15. D. A. Hejhal, *The Selberg trace formula for* PSL(2,**R**), Vol. I, Lecture Notes in Math. **548**, Springer, Berlin, 1976.

16. D. A. Hejhal, *The Selberg trace formula and the Riemann zeta function*, Duke Math. J. **43** (1976), 441–482.

17. E. Hopf, *Fuchsian groups and ergodic theory*, Trans. Am. Math. Soc. **39** (1936), 299–314.

18. A. E. Ingham, *The Distribution of Prime Numbers*, Cambridge Mathematical Tracts No. 30, Cambridge, 1932.

19. A. B. Katok, Ya. G. Sinai and A. M. Stepin, *Theory of dynamical systems and general transformation groups with invariant measure*, J. Sov. Math. **7** (1977), 974–1065.

20. J. Lehner, *A Short Course in Automorphic Functions*, Holt, Rinehart and Winston, New York, 1965.

21. A. M. Macbeath, *Discontinuous groups and birational transformations*, Proc. Summer School, Queen's Coll., Dundee, 1961.

22. D. H. Mayer, *On a zeta function related to the continued fraction transformation*, Bull. Soc. Math. France **104** (1976), 195–203.

23. W. Parry, *Bowen's equidistribution theory and the Dirichlet Density Theorem*, Ergodic Theory and Dynamic Systems **4** (1984), 117–134.

24. W. Parry and M. Pollicott, *An analogue of the prime number theorem for closed orbits of Axiom A flows*, Ann. of Math., to appear.

25. M. Pollicott, *A complex Ruelle–Perron–Frobenius theorem and two counter examples*, Ergodic Theory and Dynamic Systems **4** (1984), 135–146.

26. C. Series, *Symbolic dynamics for geodesic flows*, Acta Math. **146** (1981), 103–128.

27. C. Series, *On coding geodesics with continued fractions*, in *Ergodic Theory* (Sem. Plans-sur-Box), Monograph Enseign Math., 29, Univ. Geneve, Geneva, 1981.

28. C. Series, *The infinite word problem and limit sets in Fuchsian groups*, Ergodic Theory and Dynamic Systems **1** (1981), 337–360.

29. C. Series, *The modular surface and continued fractions*, J. London Math. Soc. (1984).

30. A. B. Venkov, *Spectral theory of automorphic functions, the Selberg zeta-function, and some problems of analytic number theory and mathematical physics*, Russ. Math. Surv. **34** (3) (1979), 79–153.

31. N. Wiener, *The Fourier Integral and Certain of its Applications*, Dover, New York, 1958.