Vol. 58, No. 3 DUKE MATHEMATICAL JOURNAL © June 1989

CLOSED GEODESICS IN HOMOLOGY CLASSES ON SURFACES OF VARIABLE NEGATIVE CURVATURE

STEVEN P. LALLEY

0. Introduction. Let M be a compact Riemannian manifold of negative curvature. It is well known that there exist countably many closed geodesics in M, and that if $N(t)$ is the number of such closed geodesics with lengths $\leqslant t$ then

$$N(t) \sim e^{ht}/ht \tag{0.1}$$

as $t \to \infty$, where $h > 0$ is the topological entropy of the flow [9]. Recently, Phillips and Sarnak [12] and Katsuda and Sunada [6] have investigated the asymptotic behavior (as $t \to \infty$) of $N(t; m)$, the number of closed geodesics in the homology class m with lengths $\leqslant t$. For manifolds M with *constant* negative curvature they prove that for each homology class m, as $t \to \infty$

$$N(t; m) \sim Ce^{ht}/t^{1+r/2} \tag{0.2}$$

for a constant $C > 0$, where r is the rank (over $\mathbb{Z}$) of the homology group $H_1 M$ (i.e., $H_1 M \cong \mathbb{Z}^r \oplus G$, where G is the torsion subgroup).

The purpose of this paper is to extend (0.2) to manifolds of *variable* negative curvature, and to describe the asymptotics of $N(t; m)$ when m varies with t in a roughly linear fashion. For simplicity we shall only consider surfaces M whose first homology groups are torsion free, i.e., $H_1 M \cong \mathbb{Z}^{2g}$, $g \geqslant 2$. There exist C^∞ forms $\omega_1, \ldots, \omega_{2g}$ on M such that for any smooth closed curve γ on M the homology class of γ is $(\int_\gamma \omega_1, \ldots, \int_\gamma \omega_{2g})$. Let SM be the unit tangent bundle of M; define W_i: $SM \to \mathbb{R}$ by $W_i(x, v) = \langle \omega_i(x), v \rangle$ (here $<, >$ denotes dot product). For $\xi \in \mathbb{R}^{2g}$ define $-\Gamma(\xi)$ to be the maximum entropy of an invariant probability measure λ on SM satisfying $\int W_i \, d\lambda = \xi_i \forall i = 1, 2, \ldots, 2g$ (invariant means invariant with respect to the geodesic flow on SM). In sec. 4 we will show that $-\Gamma(\xi)$ is well defined and C^∞ for ξ in some neighborhood of the origin, and that the Hessian matrix $\nabla^2 \Gamma(\xi)$ is strictly positive definite for every ξ in this neighborhood.

The main result of this paper is

THEOREM 1. *Let* $\xi = t^{-1}(m_1, \ldots, m_{2g})$; *then as* $t \to \infty$

$$N(t; m) \sim e^{-t\Gamma(\xi)} t^{-g-1} (2\pi)^{-g} (\det \nabla^2 \Gamma(\xi))^{1/2} (\langle \nabla \Gamma(\xi), \xi \rangle - \Gamma(\xi))^{-1} \tag{0.3}$$

uniformly for ξ *in some neighborhood of the origin.*

Received June 20, 1986. Revision received October 22, 1988.

The result (0.2) is a special case of this because $-\Gamma(0) = h$ (Prop. 7 below). Observe that the asymptotics of $N(t; m)$ for m varying linearly with $t^{1/2}$ may be deduced from (0.3); in particular, if $\zeta = t^{-1/2}(m_1, \ldots, m_{2g})$ then uniformly for ζ in any compact subset of $\mathbb{R}^{2g}$, as $t \to \infty$

$$N(t; m) \sim e^{th} \exp\{-\langle \zeta, \nabla^2\Gamma(0)\zeta\rangle/2\} t^{-g-1}(2\pi)^{-g}(\det \nabla^2\Gamma(0))^{1/2}/h. \tag{0.4}$$

(Note that $\nabla\Gamma(0) = 0$ because $-\Gamma(\xi)$ has its maximum at $\xi = 0$.) The formula (0.4) may be interpreted as a (local) "central limit theorem" for closed geodesics: if one randomly chooses a closed geodesic from the set of closed geodesics with lengths $\leqslant t$ then the distribution of the (renormalized) homology class $t^{-1/2}m$ is approximately the $2g$-dimensional Gaussian distribution with mean vector 0 and covariance matrix $\nabla^2\Gamma(0)$. Similarly, (0.3) may be interpreted as a "large deviation theorem".

Our approach to Th. 1 is completely different from that of [6] and [12], which is based on the Selberg trace formula. We use the symbolic dynamics for geodesic flows developed by Sinai, Ratner, and Bowen to reformulate the problem as a counting problem in a sequence space, then use certain aspects of Ruelle's "thermodynamic formalism" to solve this counting problem. The method is more intricate than that of [6] and [12], and the details of the Fourier analysis more demanding; moreover, it appears to be ill suited for asymptotic expansions. However, it applies to a large class of flows (those admitting "symbolic dynamics") [7], and variations on the method are suitable for a large variety of counting problems in hyperbolic geometry (see [8] for some examples).

The overall organization of the calculation is virtually the same as that in our earlier paper [7]. However, some simplifications are possible here ((i) there are no "rapidly oscillating terms", as in [7], Lemma 7; and (ii) all of the functions except the height function $r(x)$ are integer-valued, which makes the Fourier analysis less complicated). Also, there is a (correctable) error in the unsmoothing argument of [7], sec. 6. For these reasons we shall give a complete proof of Th. 1 here, without reference to [7].

Another advantage of the methods used here and in [7], [8] is that they yield as a by-product sharp information about the distribution of individual closed geodesics in SM. Let γ be a closed geodesic (considered as a path in SM) and let G: $SM \to \mathbb{R}$ be a continuous function; define $\gamma(G)$ to be the integral of G over γ, i.e., $\gamma(G) = \int_0^{|\gamma|} G(\gamma(s))\, ds$ where γ is parametrized by arclength s. Let $\bar{\mu}_\xi$ be the invariant probability measure for the geodesic flow with maximum entropy subject to the constraints $\int W_i\, d\bar{\mu}_\xi = \xi_i$, $i = 1, \ldots, 2g$. For $\varepsilon > 0$ define $N(t; m; G; \varepsilon)$ to be the number of closed geodesics γ of length $|\gamma| \leqslant t$, in homology class m, and satisfying

$$\left| \frac{\gamma(G)}{|\gamma|} - \int G\, d\bar{\mu}_\xi \right| \leqslant \varepsilon$$

where $\xi = m/t$.

THEOREM 2. *As* $t \to \infty$, $N(t; m; G; \varepsilon)/N(t; m) \to 1$ *for any* $\varepsilon > 0$, *uniformly for* $\xi = m/t$ *in some neighborhood of the origin.*

In other words, most of the closed geodesics counted in $N(t; m)$ are distributed on SM approximately (in the weak topology) as $\bar{\mu}_\xi$. We shall not prove Th. 2, as it is very similar to results proved in [7], [8] (cf. [7], Th. 4; [8], Th. 7).

Note. After writing this paper I received preprints of Pollicott and Katsuda/Sunada, each extending (0.2) to compact manifolds with variable negative curvature. The methods used in these papers do not appear to be capable of yielding the stronger result (0.3).

1. Symbolic dynamics. The counting arguments of this paper rely on the representation of the geodesic flow by a suspension flow over a shift of finite type. Geodesics are coded in a more-or-less bijective fashion into infinite sequences from a finite alphabet, with the closed geodesics corresponding to periodic sequences. The enumeration of periodic sequences then proceeds by way of "thermodynamic formalism".

Let A be an irreducible, aperiodic, $l \times l$ matrix of zeros and ones; define

$$\Sigma_A = \left\{ x \in \prod_{-\infty}^{\infty} \{1, 2, \ldots, l\} : A(x_n, x_{n+1}) = 1 \ \forall n \right\},$$

$$\Sigma_A^+ = \left\{ x \in \prod_{0}^{\infty} \{1, 2, \ldots, l\} : A(x_n, x_{n+1}) = 1 \ \forall n \right\}.$$

These spaces are compact, metrizable, and totally disconnected in the topology of coordinatewise convergence. The maps $\sigma: \Sigma_A^+ \to \Sigma_A^+$ and $\sigma: \Sigma_A \to \Sigma_A$ defined by $(\sigma x)_n = x_{n+1}$ are called the *one-sided shift* and the *two-sided shift*, respectively; the two-sided shift is a homeomorphism, whereas the one-sided shift is continuous and surjective but not injective. For each $\rho \in (0, 1)$ there are metrics d_ρ^+, d_ρ on Σ_A^+, Σ_A defined by

$$d_\rho^+(x, y) = \sum_{n=0}^{\infty} 1\{x_n \neq y_n\}/\rho^n,$$

$$d_\rho(x, y) = \sum_{-\infty}^{\infty} 1\{x_n \neq y_n\}/\rho^{|n|},$$

each inducing the topology of coordinatewise convergence. Let $\mathscr{F}_\rho^+$, $\mathscr{F}_\rho$ denote the spaces of complex-valued, Lipschitz-continuous functions on Σ_A^+, Σ_A, respectively, relative to the metrics d_ρ^+, d_ρ. Observe that $\mathscr{F}_\rho^+$ is naturally embedded in $\mathscr{F}_\rho$, and that $\mathscr{F}_\rho^+$, $\mathscr{F}_\rho$ are Banach spaces when endowed with the norms $\|\cdot\|_\rho = \|\cdot\|_\infty + |\cdot|_\rho$, where $|f|_\rho = \sup\{|f(x) - f(y)|/d_\rho(x, y)\}$ (or d_ρ^+).

Suspension flows over the shift (Σ_A, σ) are defined as follows. Let $r \in \mathscr{F}_\rho$ be a strictly positive function on Σ_A (r is called the *height function*), and let

$$\Sigma_A^r = \{(x, t): x \in \Sigma_A \text{ and } 0 \leqslant t \leqslant r(x)\}$$

with the points $(x, r(x))$ and $(\sigma x, 0)$ identified for each $x \in \Sigma_A$. The suspension flow σ_t^r, $-\infty < t < \infty$, on Σ_A^r is defined by

$$\sigma_t^r(x, s) = (x, s + t) \qquad \text{if} \quad 0 \leqslant s \leqslant r(x) \text{ and } 0 \leqslant s + t \leqslant r(x),$$

$$\sigma_t^r \circ \sigma_s^r = \sigma_{t+s}^r \qquad \forall s, t \in \mathbb{R}.$$

The dynamics of the flow σ_t^r may be described as follows: starting at any point $(x, s) \in \Sigma_A^r$, move at unit speed up the vertical fiber over $(x, 0)$ until reaching $(x, r(x))$, then jump instantaneously to $(\sigma x, 0)$ and continue along the vertical fiber over $(\sigma x, 0)$. Observe that the periodic orbits of this flow are precisely those orbits that intersect the "floor" $\Sigma_A \times \{0\}$ at points $(x, 0)$ where x is a periodic sequence. If an orbit passes through $(x, 0)$ where $\sigma^n x = x$ and $\sigma^i x \neq x$ for $i = 1, 2, \ldots, n - 1$, then the orbit is periodic with least period

$$S_n r(x) \triangleq r(x) + r(\sigma x) + \cdots + r(\sigma^{n-1} x).$$

Consider now the geodesic flow Φ on the unit tangent bundle SM of a compact, C^∞, Riemannian manifold M. This flow is known to be an Anosov flow, hence the results of Bowen [3] and Ratner [17] are applicable. In particular, there exists a suspension flow σ^r on Σ_A^r, with height function $r \in \mathscr{F}_\rho$ for *some* $0 < \rho < 1$, and a Lipschitz continuous map $\pi: \Sigma_A^r \to SM$ such that

(1.1) *π is surjective;*
(1.2) *π is at most N to 1 for some $N < \infty$;*
(1.3) *$\pi \circ \sigma_t^r = \Phi_t \circ \pi \; \forall t \geqslant 0$; and*
(1.4) *all but finitely many of the periodic orbits $\{\gamma\}$ of Φ have the property that $\pi^{-1}(\gamma)$ consists of a single periodic orbit of σ^r with the same least period.*

Because the representation of the geodesic flow Φ as a suspension flow is fundamental to our analysis, we shall give a resumé of the main features of the Bowen/Ratner construction. Codimension one "rectangles" $R_1, R_2, \ldots, R_l$ are constructed in SM transverse to the flow Φ; each side of R_i lies either in a leaf of the stable foliation W^s or in a leaf of the unstable foliation W^u. Now each orbit of Φ cuts through $\bigcup_{i=1}^l R_i$ in a doubly infinite sequence of points, and if the rectangles R_i are suitably chosen then the sequence of *indices* $\ldots i_{-1} i_0 i_1 \ldots$ uniquely determines the orbit, and conversely for a suitable transition matrix A every sequence in Σ_A corresponds to an orbit. The assumption of negative curvature is crucial for this because it ensures that no two orbits pass through the same sequence of rectangles R_i. The correspondence between orbits and sequences is $1 - 1$ *except* for orbits

which pass through the boundary of some R_i. (Note: only finitely many *closed* orbits of Φ pass through the boundary of some R_i, since these boundaries lie in leaves of W^u or W^s; this explains (1.4) above.)

Assume now that M is a compact surface of genus $g \geqslant 2$. Let $W_1, W_2, \ldots, W_{2g}: SM \to \mathbb{R}$ be as in sec. 0; thus for any smooth closed curve $\gamma(s)$ parametrized by arc-length the homology class of γ in $H_1 M \cong \mathbb{Z}^{2g}$ is

$$\left(\int_0^{|\gamma|} W_1(\gamma(s), \gamma'(s))\, ds, \ldots, \int_0^{|\gamma|} W_{2g}(\gamma(s), \gamma'(s))\, ds \right).$$

For $x \in \Sigma_A$ the path (x, t), $0 \leqslant t \leqslant r(x)$ projects via π to a segment of an orbit of the geodesic flow in SM. Define

$$\varphi_i(x) = \int_{t=0}^{r(x)} W_i(\pi(x, t))\, dt, \; i = 1, \ldots, 2g.$$

Let $x \in \Sigma_A$ be a periodic sequence with smallest period n; then the path (x, t), $0 \leqslant t \leqslant S_n r(x)$, is a periodic orbit of the suspension flow with least period $S_n r(x)$. The projection via π of this periodic orbit is a closed geodesic. If the minimal period of this closed geodesic is also $S_n r(x)$ then its homology class is $(S_n \varphi_1(x), \ldots, S_n \varphi_{2g}(x))$ where $S_n f = f + f \circ \sigma + \cdots + f \circ \sigma^{n-1}$. It therefore follows from (1.4) above that the number of closed geodesics in M with least period $\leqslant \tau$ and in homology class $(m_1, m_2, \ldots, m_{2g})$ is $R(\tau; m_1, \ldots, m_{2g}) + 0(1)$, where

$$(1.5) \quad R(\tau; m_1, \ldots, m_{2g}) \triangleq \sum_{n=1}^{\infty} n^{-1} \sum_{x \in \mathscr{P}_n} 1\{S_n r(x) \leqslant \tau;\, S_n \varphi_i(x) = m_i \; \forall i = 1, \ldots, 2g\}$$

and $\mathscr{P}_n$ is the set of periodic sequences in Σ_A with least period n. In the subsequent sections we shall undertake an asymptotic analysis of the function $R(\tau; m_1, \ldots, m_{2g})$.

The symbolic dynamics described in (a)–(g) above is by no means canonical (in fact, Bowen's construction shows that there are infinitely many such representations). In the remainder of this section we will show that the suspension flow may be chosen in such a way that the functions $r, \varphi_1, \ldots, \varphi_{2g}$ are in a form advantageous to the Fourier analysis of secs. 2–3 below.

Two functions $f, g \in \mathscr{F}_\rho$ are said to be *cohomologous* if there exists a continuous function h on Σ_A such that $f - g = h - h \circ \sigma$. Note that if f, g are cohomologous and if $\sigma^n x = x$ then $S_n f(x) = S_n g(x)$. According to Lemma 1.6 of [4], for any $f \in \mathscr{F}_\rho$ there exists $f^* \in \mathscr{F}^+_{\sqrt{\rho}}$ such that f and f^* are cohomologous; a close examination of the proof shows that if $f > 0$ then f^* can be chosen so that $f^* > 0$. Let r^*, $\varphi_i^* \in \mathscr{F}^+_{\sqrt{\rho}}$ be such that r, r^* are cohomologous and $r^* > 0$, and φ_i, φ_i^* are cohomologous. If $x \in \Sigma_A$ satisfies $\sigma^n x = x$ then $S_n r(x) = S_n r^*(x)$ and $S_n \varphi_i(x) = S_n \varphi_i^*(x)$, so none of the quantities in (1.5) is changed if r is replaced by r^* and φ_i is replaced by φ_i^*. Moreover, Prop. 2 of [15] implies that φ_i^* may be chosen so as to be an integer-

valued function, because if $\sigma^n x = x$ then $S_n\varphi_i(x)$ is integer-valued, as it is the i^{th} coordinate of the homology class of a closed geodesic.

The main result of [16] and Prop. 7 of [14] imply that the height function r cannot be cohomologous to any piecewise constant function. Since $\varphi_1^*, \ldots, \varphi_{2g}^*$ are integer-valued, they are piecewise constant; hence

$$r \neq \sum_{j=1}^{2g} a_j\varphi_j^* + \psi + h - h \circ \sigma \tag{1.6}$$

for any scalars a_j, any piecewise constant ψ, and any continuous h.

LEMMA 0. *The vector-valued function $(\varphi_1^*, \ldots, \varphi_{2g}^*)$ is not cohomologous to any function valued in a proper subgroup of $\mathbb{Z}^{2g}$.*

Proof. Each homology class $m = (m_1, \ldots, m_{2g}) \in \mathbb{Z}^{2g} \cong H_1 M$ contains a closed geodesic ([2], sec. 11.7, Th. 10). Hence, by (1.4), there exist closed geodesics $\gamma_1, \ldots, \gamma_{2g}$ in homology classes $m^1, \ldots, m^{2g}$ such that $m^1, \ldots, m^{2g}$ generate $\mathbb{Z}^{2g}$ and such that the preimage (under π) of any γ_i consists of a single periodic orbit of the suspension flow with the same minimal period. It follows that there are periodic sequences $x^1, x^2, \ldots, x^{2g} \in \Sigma_A$ with least periods $n_1, \ldots, n_{2g}$, such that $S_{n_i}\varphi^*(x^i) = m^i$ for each $i = 1, \ldots, 2g$. Since $m^1, \ldots, m^{2g}$ generate $\mathbb{Z}^{2g}$, this proves that φ^* cannot be cohomologous to a function valued in a proper subgroup of $\mathbb{Z}^{2g}$. □

PROPOSITION 0. *The suspension flow may be chosen so that $(\varphi_1^*, \ldots, \varphi_{2g}^*)$ is not cohomologous to any function valued in a coset of a proper subgroup of $\mathbb{Z}^{2g}$.*

The proof is given in the Appendix. It is not absolutely essential for the Fourier analysis of sects. 2–3 that $(\varphi_1^*, \ldots, \varphi_{2g}^*)$ have this property; the important fact is really Lemma 0. However, it does simplify the calculations somewhat. As to the proof of Prop. 0, we remark that it is based on a coding trick—no new geometric argument is needed. We just change the sequence space by replacing each symbol $i \in \{1, 2, \ldots, l\}$ by a block $i_1 i_2 \ldots i_{M(i)}$ and redefining the functions r, r^*, φ_j, etc. accordingly. By choosing $M(1), M(2), \ldots, M(l)$ appropriately we get rid of unwanted "periodicities" in $\varphi_1^*, \ldots, \varphi_{2g}^*$.

In the remaining sections of the paper (excluding the appendix) we shall drop the superscript * on $r^*, \varphi_1^*, \ldots, \varphi_{2g}^*$, as the original $r, \varphi_1, \ldots, \varphi_{2g}$ will play no further role. Thus, from here on $r, \varphi_1, \ldots, \varphi_{2g}$ are functions in $\mathscr{F}_\rho^+$ such that $(r, \varphi_1, \ldots, \varphi_{2g})$ is valued in $(0, \infty) \times \mathbb{Z}^{2g}$ and $(\varphi_1, \ldots, \varphi_{2g})$ is not cohomologous to any function valued in a coset of a proper subgroup of $\mathbb{Z}^{2g}$. Note that (1.5) is still valid.

2. Ruelle operators and thermodynamic functions for the shift. For any function $f \in \mathscr{F}_\rho^+$ the Ruelle operator $\mathscr{L}_f: \mathscr{F}_\rho^+ \to \mathscr{F}_\rho^+$ is defined by

$$\mathscr{L}_f g(x) = \sum_{y:\sigma y = x} e^{f(y)} g(y).$$

This is a bounded linear operator; if f is real-valued it is a positive operator. The spectrum of $\mathscr{L}_f$ has been described by Ruelle [18] and Pollicott [15]:

(a) *If f is real-valued then there is a simple eigenvalue $\lambda_f \in (0, \infty)$ whose eigenfunction $h_f \in \mathscr{F}_\rho^+$ is strictly positive on Σ_A^+. The rest of the spectrum is contained in a disc of radius $< \lambda_f$. Moreover, there is a positive Borel measure ν_f on Σ_A^+ such that $\mathscr{L}_f^* \nu_f = \lambda_f \nu_f$.*

(b) *If $f = u + iv$, where $u, v \in \mathscr{F}_\rho^+$ are real-valued, then (b_1) if for some constant $a \in [-\pi, \pi]$ the function $(v - a)/2\pi$ is cohomologous to an integer-valued function then $e^{ia}\lambda_u$ is a simple eigenvalue of $\mathscr{L}_f$, and the rest of the spectrum is contained in a disc of radius $< \lambda_u$; (b_2) otherwise, the entire spectrum of $\mathscr{L}_f$ is contained in a disc of radius $< \lambda_u$.*

We may assume that the eigenfunctions and eigenmeasures in (a) are normalized so that

$$1 = \int 1 \, d\nu_0 = \int h_f \, d\nu_0 = \int h_f \, d\nu_f.$$

Standard perturbation theory shows that λ_f, h_f are analytic in f in some open neighborhood of the real-valued functions in $\mathscr{F}_\rho^+$; ν_f is weakly analytic in f in the sense that for any $g \in \mathscr{F}_\rho^+$ the real-valued function $\int g \, d\nu_f$ is analytic. Observe that if f, f^* are cohomologous then $\lambda_f = \lambda_{f^*}$; also, if $f^* = f + g - g \circ \sigma$ then $h_{f^*} = e^{-g} h_f$ and $d\nu_{f^*} = e^g \, d\nu_f$.

For each real-valued $f \in \mathscr{F}_\rho^+$ the measure μ_f defined by $d\mu_f = h_f \, d\nu_f$ is a σ-invariant probability measure on Σ_A^+ called the *Gibbs state* or *equilibrium state* for f ([5], ch. 1). Because it is σ-invariant, μ_f extends to a σ-invariant measure on Σ_A which we also denote by μ_f. If f, g are cohomologous then $\mu_f = \mu_g$; otherwise μ_f, μ_g are mutually singular. The Gibbs state μ_f is a strongly mixing invariant measure for σ.

The asymptotic expansions of section 0 involve the entropies of certain invariant measures. The key to identifying the quantities in these expansions as entropies is the *Gibbs variational principle*. This states that if $f \in \mathscr{F}_\rho^+$ is real-valued and if μ is a σ-invariant Borel probability measure on Σ_A then

$$\log \lambda_f \geqslant H(\mu, \sigma) + \int f \, d\mu \tag{2.1}$$

with equality iff $\mu = \mu_f$ (here $H(\mu, \sigma)$ is the entropy of μ relative to σ).

Recall now the functions $r, \varphi_1, \ldots, \varphi_{2g} \in \mathscr{F}_\rho^+$ constructed in sec. 1. For $z = (z_0, z_1, \ldots, z_{2g}) \in \mathbb{R}^{2g+1}$ define the *pressure function*

$$\beta(z) = \log(\lambda_{z_0 r + \sum_{i=1}^{2g} z_i \varphi_i}).$$

It is known ([19], Ch. 5, Ex. 5) that

$$\partial\beta/\partial z_0 = \int r \, d\mu_z, \qquad \partial\beta/\partial z_i = \int \varphi_i \, d\mu_z, \qquad i \geqslant 1; \tag{2.2}$$

$$\nabla^2\beta(z) \text{ is strictly positive definite } \forall z \in \mathbb{R}^{2g+1}, \tag{2.3}$$

where $\mu_z = \mu_{z_0 r + \Sigma z_i \varphi_i}$ is the Gibbs state associated with the function $z_0 r + \sum_{i=1}^{2g} z_i \varphi_i$. (The fact that $\nabla^2\beta(z)$ is *strictly* positive definite uses the fact that no linear combination of $r, \varphi_1, \ldots, \varphi_{2g}$ is cohomologous to a constant; cf. (1.6).) It follows that β is strictly convex, strictly increasing in z_0, and that $\nabla\beta$ is a diffeomorphism of $\mathbb{R}^{2g+1}$ onto an open subset Ω of $\mathbb{R}^{2g+1}$.

The Legendre transform γ of β is defined by

$$\gamma(\xi) = \sup_{z \in \mathbb{R}^{2g+1}} (\langle \xi, z \rangle - \beta(z)), \ \xi \in \mathbb{R}^{2g+1},$$

where $<, >$ denotes dot product. The function γ is convex on $\mathbb{R}^{2g+1}$. For $\xi \in \Omega$ the sup is attained uniquely at that z for which $\nabla\beta(z) = \xi$, because of the strict convexity of β. The inverse function theorem therefore implies that

$$\nabla\gamma \circ \nabla\beta = \text{identity on } \mathbb{R}^{2g+1}; \tag{2.4}$$

$$\nabla\beta \circ \nabla\gamma = \text{identity on } \Omega; \tag{2.5}$$

$$\nabla^2\gamma(\xi) = (\nabla^2\beta(z))^{-1} \qquad \text{if } \nabla\beta(z) = \xi. \tag{2.6}$$

Thus γ is strictly convex on Ω. The Gibbs variational principle implies that

$$\gamma(\xi) = -H(\mu_z, \sigma) \qquad \text{if } \nabla\beta(z) = \xi.$$

For $\zeta = (\zeta_0, \zeta_1, \ldots, \zeta_{2g}) \in \mathbb{C}^{2g+1}$ we shall use the abbreviated notation

$$\mathscr{L}_\zeta = \mathscr{L}_{\zeta_0 r + \sum_{j=1}^{2g} \zeta_j \varphi_j}.$$

PROPOSITION 1. *For each $z \in \mathbb{R}^{2g+1}$ there is an open neighborhood $\mathscr{N}$ of z in $\mathbb{C}^{2g+1}$ and an $\varepsilon > 0$ such that for every $\zeta \in \mathscr{N}$ and every $g \in \mathscr{F}_\rho^+$*

$$\left\| e^{-n\beta(\zeta)} \mathscr{L}_\zeta^n g - \left(\int g \, d\nu_\zeta \right) h_\zeta \right\|_\rho \leqslant C \|g\|_\rho (1 + \varepsilon)^{-n} \ \forall n \geqslant 1. \tag{2.7}$$

Proof. Recall that the functions $\zeta \to h_\zeta, \nu_\zeta, \beta(\zeta)$ extend to analytic functions for ζ in some neighborhood of z. In this neighborhood,

$$\mathscr{L}_\zeta = \lambda_\zeta \mathscr{L}_\zeta' + \mathscr{L}_\zeta'' \text{ where } \mathscr{L}_\zeta' g = \left(\int g \, d\nu_\zeta \right) h_\zeta.$$

Clearly, spectrum $(\mathscr{L}_\zeta'') = (\text{spectrum } (\mathscr{L}_\zeta)\backslash\{\lambda_\zeta\}) \cup \{0\}$; thus for all ζ near z the spectral radius of $\mathscr{L}_\zeta''$ is $\leqslant \lambda_z - \delta$ for some $\delta > 0$. Moreover, for any $n \geqslant 1$

$$\mathscr{L}_\zeta^n = (\lambda_\zeta \mathscr{L}_\zeta')^n + (\mathscr{L}_\zeta'')^n$$

$$= \lambda_\zeta^n \mathscr{L}_\zeta' + (\mathscr{L}_\zeta'')^n;$$

since $\beta(\zeta) = \log \lambda_\zeta$, (2.7) follows from the spectral radius formula. □

PROPOSITION 2. *For any $z \in \mathbb{R}^{2g+1}$ and small $\delta > 0$ there exist $\varepsilon > 0$, $C < \infty$ such that if $|\theta_0| \leqslant \delta^{-1}$ and $\delta \leqslant |\theta_j| \leqslant \pi$ for $j = 1, \ldots, 2g$ then*

$$\|\mathscr{L}_{z+i\theta}^n\| \leqslant Ce^{n\beta(z)-n\varepsilon} \ \forall n \geqslant 1 \tag{2.8}$$

where $\theta = (\theta_0, \theta_1, \ldots, \theta_{2g})$. Moreover, C and ε can be chosen so as to vary continuously with z.

Proof. Recall (sec. 1) that the vector-valued function $(\varphi_1, \ldots, \varphi_{2g})$ is not cohomologous to any function valued in a coset of a proper subgroup of $\mathbb{Z}^{2g}$ and that r is not cohomologous to any piecewise constant function. By Pollicott's theorem (statement (b) at the beginning of this section), the spectral radius of $\mathscr{L}_\zeta$ is strictly less than λ_z unless $\theta_0 r + \Sigma_{j=1}^{2g} \theta_j \varphi_j$ is cohomologous to a constant plus a function valued in $2\pi\mathbb{Z}$. This is impossible unless $\theta_0 = 0$ and $(\theta_1, \ldots, \theta_{2g}) \in 2\pi\mathbb{Z}^{2g}$. Since λ_ζ and spectral radius $(\mathscr{L}_\zeta)^{2g}$ are continuous functions near $z \in \mathbb{R}^{2g+1}$, (2.8) follows from the fact that $\{\theta: |\theta_0| \leqslant \delta^{-1}$ and $\delta \leqslant |\theta_j| \leqslant \pi \ \forall j = 1, \ldots, 2g\}$ is compact, by the spectral radius formula. □

3. A saddlepoint calculation. Our approach to (0.3) will be to analyze the series (1.5) term by term, using Fourier analysis. For various reasons it is easier to work with modified versions of these terms; thus, for $t, \tau \in (0, \infty)$ and $m_1, \ldots, m_{2g} \in \mathbb{Z}$ define

$$Q_n(t, m; \tau) = \sum_{x: \sigma^n x = x} 1\{0 \leqslant S_n r(x) - t \leqslant \tau; S_n \varphi_i(x) = m_i \ \forall 1 \leqslant i \leqslant 2g\}.$$

PROPOSITION 3. *Let $n^{-1}(t, m_1, \ldots, m_{2g}) = (\xi_0, \xi_1, \ldots, \xi_{2g}) = \xi$. If $\xi \in \Omega$ and $\xi = \nabla\beta(z)$ for some $z \in \mathbb{R}^{2g+1}$ then*

$$Q_n(t, m; \tau) \sim e^{-n\gamma(\xi)}(2\pi n)^{-(2g+1)/2}(\det \nabla^2\gamma(\xi))^{1/2} \int_0^\tau e^{-z_0 s}\, ds \tag{3.1}$$

as $n \to \infty$, uniformly for ξ in any compact subset of Ω.

PROPOSITION 4. *There exist positive constants K_z varying continuously with $z \in \mathbb{R}^{2g+1}$ such that for each z, t, m,*

$$Q_n(t, m; \tau) \leqslant K_z \exp\{n\beta(z) - \langle z, \zeta\rangle + |z_0|\tau\} \tag{3.2}$$

for every $n \geqslant 1$, where $\zeta = (t, m_1, m_2, \ldots, m_{2g})^t$.

The rest of this section is devoted to the proofs of these results. To use Fourier analysis we must relate the Fourier transform of Q_n to the thermodynamic functions β, γ; for this we shall appeal to Ruelle's operator theorem. Thus, the first order of business is to replace the sum $\Sigma_{x:\sigma^n x=x}$ by sums $\Sigma_{x:\sigma^n x=y}$.

Fix k; let $x^{(1)}, x^{(2)}, \ldots, x^{(p)} \in \Sigma_A^+$ be chosen in such a way that the set $\{x_0^{(i)}x_1^{(i)}\ldots x_k^{(i)}, i = 1, 2, \ldots, p\}$ is the set of all *finite* sequences of length $k+1$ from the alphabet $\{1, 2, \ldots, l\}$ with transitions allowed by A. Assume that no two of the finite sequences $x_0^{(i)}x_1^{(i)}\ldots x_k^{(i)}$ are the same. Define

$$g_i(x) = 1\{x_n = x_n^{(i)}\ \forall 0 \leqslant n \leqslant k\},$$

$$Q_n^i(t, m; \tau) = \sum_{x:\sigma^n x = x^{(i)}} g_i(x) 1\{0 \leqslant S_n r(x) - t \leqslant \tau;\ S_n\varphi_j(x) = m_j\ \forall 1 \leqslant j \leqslant 2g\}.$$

Observe that $\Sigma_{i=1}^p g_i \equiv 1$. Consider x such that $\sigma^n x = x$, $n > k$, and suppose $g_i(x) = 1$. The sequence $\tilde{x}$ defined by $\tilde{x}_j = x_j$, $1 \leqslant j \leqslant n+k$, $\tilde{x}_{j+n} = x_j^{(i)}$ has the property that $\sigma^n\tilde{x} = x^{(i)}$ and $g_i(\tilde{x}) = 1$, hence is included in the sum defining Q_n^i; moreover, $d_\rho(x, \tilde{x}) \leqslant (\text{const})\rho^{n+k}$, so $|S_n r(x) - S_n r(\tilde{x})| \leqslant (\text{const})\rho^k$ and $S_n\varphi_j(x) = S_n\varphi_j(\tilde{x})$ (if k is sufficiently large). Consequently,

$$\sum_{i=1}^p Q_n^i(t, m; \tau) \leqslant Q_n(t - \varepsilon_k, m; \tau + 2\varepsilon_k), \tag{3.3}$$

$$Q_n(t, m; \tau) \leqslant \sum_{i=1}^p Q_n^i(t - \varepsilon_k, m; \tau + 2\varepsilon_k), \tag{3.4}$$

where $\varepsilon_k = (\text{const})\rho^k$. Therefore, to obtain (3.1)–(3.2) it will suffice to analyze Q_n^i, since by choosing k large we can make ε_k arbitrarily small.

Proof of Proposition 4. Fix $z \in \mathbb{R}^{2g+1}$; then

$$Q_n^i(t, m; \tau)\exp\left\{\sum_{j=0}^{2g} z_j\zeta_j\right\}\exp\{-|z_0|\tau\} \leqslant \sum_{\sigma^n x = x^{(i)}} \exp\left\{z_0 S_n r(x) + \sum_{j=1}^{2g} z_j S_n\varphi_j(x)\right\} g_i(x)$$

$$= \mathscr{L}_z^n g_i(x^{(i)}) \leqslant K_z\lambda_z^n = K_z e^{n\beta(z)}$$

by the spectral radius formula and Ruelle's theorem. The inequality (3.2) now follows by an easy argument from (3.4). The continuity in z of K_z follows from the continuity of $z \to \mathscr{L}_z$. □

To prove (3.1) we will show that for each $i = 1, 2, \ldots, p$, as $n \to \infty$

$$Q_n^i(t, m; \tau) \sim C_i(z)e^{-n\gamma(\xi)}(2\pi n)^{-(2g+1)/2}(\det \nabla^2\gamma(\xi))^{1/2}\int_0^\tau e^{-z_0 s}\, ds \tag{3.5}$$

uniformly for ξ in any compact subset of Ω, where $C_i(z) = (\int g_i\, dv_z)h_z(x^{(i)})$ (here v_z, h_z are the eigenmeasure and eigenfunction for the leading eigenvalue λ_z of $\mathscr{L}_z$). Recall that $x_0^{(i)} \dots x_k^{(i)}$, $i = 1, 2, \dots, p$ are the distinct sequences of length $k+1$ with transitions allowed by A, and that $g_i(x) = 1\{x_n = x_n^{(i)}\ \forall 0 \leqslant n \leqslant k\}$. If k is large then

$$\sum_{i=1}^{p} C_i(z) \approx \int h_z\, dv_z = 1,$$

since $h_z(x)$ is a continuous function of x. Consequently, in view of (3.3)–(3.4), (3.5) implies (3.1).

For $n \geqslant 1$, $i \in \{1, 2, \dots, p\}$, and $m = (m_1, \dots, m_{2g}\} \in \mathbb{Z}^{2g}$ define a positive Borel measure $M_{n,m}^i(ds)$ on $\mathbb{R}$ by

$$M_{n,m}^i(ds) = \sum_{\sigma^n x = x^{(i)}} g_i(x) 1\{S_n r(x) \in ds;\ S_n \varphi_j(x) = m_j\ \forall 1 \leqslant j \leqslant 2g\}.$$

Observe that $Q_n^i(t, m; \tau) = \int_t^{t+\tau} M_{n,m}^i(ds)$. A standard approximation argument shows that to prove (3.5) it suffices to prove that for every nonnegative, compactly supported, C^∞ function $u\colon \mathbb{R} \to \mathbb{R}$

$$\int u(s-t) M_{n,m}^i(ds) \sim C_i(z) e^{-n\gamma(\xi)} (2\pi n)^{-(2g+1)/2} (\det \nabla^2 \gamma(\xi))^{1/2} \int u(s) e^{-z_0 s}\, ds \tag{3.6}$$

uniformly for ξ in any compact subset of Ω. Define yet another positive Borel measure $N_{n,m}^i(ds)$ on $\mathbb{R}$ by

$$N_{n,m}^i(ds) = e^{z_0 s} M_{n,m}^i(ds);$$

then to prove (3.6) it suffices to prove that for every compactly supported, C^∞ function $u \geqslant 0$,

$$\int u(s-t) N_{n,m}^i(ds) \sim C_i(z) e^{-n\gamma(\xi) + z_0 t} (2\pi n)^{-(2g+1)/2} (\det \nabla^2 \gamma(\xi))^{1/2} \int u(s)\, ds. \tag{3.7}$$

This we will accomplish by Fourier analysis.

Unfortunately, we have no control over the behavior of the Fourier-Stieltjes transform of $N_{n,m}^i$ at $i\infty$, so an additional unsmoothing argument is necessary. (This is also the reason for making the transformation from $M_{n,m}^i$ to $N_{n,m}^i$.) Let $\mathscr{P}$ be the set of even probability density functions $k(s)$, $s \in \mathbb{R}$, whose Fourier transforms $\hat{k}(i\theta) = \int e^{i\theta s} k(s)\, ds$ are *compactly supported* (in $i\mathbb{R}$). It is well known that $\mathscr{P} \neq \varnothing$.

LEMMA 1. *To prove* (3.7) *it suffices to prove that for every* $k \in \mathscr{P}$

$$\int u(s-t)(k * N_{n,m}^i)(ds) \sim \text{R.H.S. (3.7)} \tag{3.8}$$

uniformly for ξ *in any compact subset of* Ω, *as* $n \to \infty$.

The proof will be given later. To analyze (3.8) we use Parseval's formula to rewrite the integral in terms of Fourier transforms:

$$\int u(s-t)(k * N^i_{n,m})(ds) = (2\pi)^{-1} \int_{-\infty}^{\infty} \hat{u}(i\theta_0)\hat{k}(-i\theta_0)\hat{N}^i_{n,m}(-i\theta_0)e^{it\theta_0}\, d\theta_0$$

$$= (2\pi)^{-1} \int_{-\infty}^{\infty} \hat{u}(i\theta_0)\hat{k}(-i\theta_0)\hat{M}^i_{n,m}(-i\theta_0 + z_0)e^{it\theta_0}\, d\theta_0$$

(here $\hat{u}(\zeta) = \int e^{\zeta s}u(s)\, ds$, $\hat{N}^i_{n,m}(\zeta) = \int e^{\zeta s}N^i_{n,m}(ds)$, etc.). Now we express $\hat{M}^i_{n,m}$ in terms of the Ruelle operators. Observe that

$$\sum_{m \in \mathbb{Z}^{2g}} \hat{M}^i_{n,m}(\zeta_0) \exp\left\{\sum_{j=1}^{2g} \zeta_j m_j\right\} = \sum_{\sigma^n x = x^{(i)}} \exp\left\{\zeta_0 S_n r(x) + \sum_{j=1}^{2g} \zeta_j S_n \varphi_j(x)\right\} g_i(x)$$

$$= \mathscr{L}^n_\zeta g_i(x^{(i)}) \ \forall \zeta \in \mathbb{C}^{2g+1}.$$

Consequently, the inversion formula for Fourier series and Fubini's theorem (since $\hat{k}$ has compact support) imply that

$$(2\pi)^{-1} \int_{-\infty}^{\infty} \hat{u}(i\theta_0)\hat{k}(-i\theta_0)\hat{M}^i_{n,m}(-i\theta_0 + z_0)e^{it\theta_0}\, d\theta_0$$

$$= (2\pi)^{-2g-1} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} \hat{u}(i\theta_0)\hat{k}(-i\theta_0)\mathscr{L}^n_{z-i\theta} g_i(x^{(i)})$$

$$\cdot \exp\left\{it\theta_0 + \sum_{j=1}^{2g} m_j(i\theta_j - z_j)\right\} d\theta_0\, d\theta_1 \ldots d\theta_{2g}$$

$$= (2\pi)^{-2g-1} \exp\left\{-n \sum_{j=0}^{2g} z_j\xi_j + z_0 t\right\} \cdot \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} \hat{u}(i\theta_0)\hat{k}(-i\theta_0)\mathscr{L}^n_{z-i\theta} g_i(x^{(i)})$$

$$\cdot \exp\left\{in \sum_{j=0}^{2g} \theta_j\xi_j\right\} d\theta_0\, d\theta_1 \ldots d\theta_{2g}.$$

For $\theta = (\theta_0, \theta_1, \ldots, \theta_{2g})$ near the origin,

$$\mathscr{L}^n_{z-i\theta} g_i(x^{(i)}) \sim \exp\{n\beta(z - i\theta)\}\left(\int g_i\, dv_{z-i\theta}\right) h_{z-i\theta}(x^{(i)}),$$

by Prop. 1, and for θ bounded away from the origin

$$\|\mathscr{L}^n_{z-i\theta} g_i\| \leqslant (\text{const}) \exp\{n\beta(z) - n\varepsilon\}$$

for some $\varepsilon > 0$, provided $\theta_0 \in$ support $(\hat{k})$, by Prop. 2. Thus the main contribution to the last integral comes from θ near the origin, say $|\theta| \leqslant \delta$ (the fact that $\hat{k}$ has compact support is crucial for this). For θ near the origin, $(\int g_i \, dv_{z-i\theta}) h_{z-i\theta}(x^{(i)}) \approx (\int g_i \, dv_z) h_z(x^{(i)}) = C_i(z)$. Consequently,

$$\int u(s-t)(k * N^i_{n,m})(ds) \sim (2\pi)^{-2g-1} \exp\left\{n\beta(z) - n \sum_{j=0}^{2g} z_j \xi_j + z_0 t\right\} C_i(z)$$

$$\cdot \int\int \cdots \int_{|\theta| \leqslant \delta} \hat{u}(i\theta)\hat{k}(-i\theta) \exp\left\{n\left(\beta(z - i\theta) - \beta(z) - i \sum_{j=0}^{2g} \theta_j \xi_j\right)\right\} d\theta_0 \, d\theta_1 \ldots d\theta_{2g}.$$

Note that $\beta(z) - \sum_{j=0}^{2g} z_j \xi_j = -\gamma(\xi)$, because $\nabla\beta(z) = \xi$, so the leading exponential factor is $e^{-n\gamma(\xi)+z_0 t}$, as desired. Also

$$\beta(z - i\theta) - \beta(z) + i \sum_{j=0}^{2g} \theta_j \xi_j = -\langle \theta, \nabla^2 \beta(z)\theta \rangle / 2 + 0(|\theta|^2)$$

as $|\theta| \to 0$. Therefore, the last integral may be evaluated by Laplace's method of asymptotic expansion, yielding

$$\int u(s-t)(k * N^i_{n,m}(ds)) \sim (2\pi n)^{-(2g+1)/2} e^{-n\gamma(\xi)+z_0 t} C_i(z) (\det \nabla^2 \beta(z))^{-1/2} \hat{u}(0)\hat{k}(0),$$

which is the same as R.H.S. (3.7). This holds uniformly for ξ in compact subsets of Ω, because all the approximations made involve only the thermodynamic functions $z \to \mathscr{L}_z$, $\beta(z)$, v_z, h_z, which vary continuously with z, and hence with $\xi = \nabla\beta(z)$. Except for the proof of Lemma 1, this completes the proof of Proposition 3.

Proof of Lemma 1. First we will show that there is a constant $C = C(\xi) < \infty$ varying continuously with ξ such that

$$\sup_{J : \varepsilon < |J| < \varepsilon^{-1}} N^i_{n,m}(J) \leqslant C|J| n^{-(2g+1)/2} e^{-n\gamma(\xi)+z_0 t}. \tag{3.9}$$

Here the sup is over all intervals J whose lengths $|J|$ are between ε and $1/\varepsilon$. Let $v(s)$ be a nonnegative, compactly supported, C^∞ function such that $v(s) = 1 \; \forall s \in J$ and $\int v(s) \, ds \leqslant 2|J|$; then $N^i_{n,m}(J) \leqslant \int v(s) N^i_{n,m}(ds)$. Let $k \in \mathscr{P}$ be such that $\int_{-\delta}^{\delta} k(s) \, ds > 1 - \delta$ for some small $\delta > 0$. Then since $v \geqslant 0$,

$$\int v(s) N^i_{n,m}(ds) \leqslant (\text{const}) \int v(s)(k * N^i_{n,m})(ds)$$

(the constant depends only on δ and on the modulus of continuity of v). But

$$\int v(s)(k * N^i_{n,m})(ds) = (2\pi)^{-(2g+1)/2} \exp\left\{-n \sum_{j=0}^{2g} z_j\xi_j + z_0 t\right\}$$

$$\cdot \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} (\hat{v}(i\theta_0) \exp\{-it\theta_0\})\hat{k}(-i\theta_0)\mathscr{L}^n_{z-i\theta} g_i(x^{(i)})$$

$$\cdot \exp\left\{in \sum_{j=0}^{2g} \theta_j\xi_j\right\} d\theta_0 \ldots d\theta_{2g}$$

as before. Using Laplace's method and Props. 1–2 as earlier one may show that

$$\int v(s)(k * N^i_{n,m})(ds) \leqslant (\text{const})\hat{v}(0)n^{-(2g+1)/2} e^{-n\gamma(\xi)+z_0 t};$$

since $\hat{v}(0) = \int v \leqslant 2|J|$, this proves (3.9). The uniformity in J follows from the fact that the corresponding functions v may all be taken from $\{v_{t_1,t_2}(s) = v((s - t_1)/t_2)$, $t_1 \in \mathbb{R}$ and $t_2 \in K\}$ where K is a compact subset of $(0, \infty)$. The fact that the constant $C = C(\xi)$ can be made to vary continuously in ξ follows again from the fact that the approximations made in using the Laplace method depend only on the continuity and smoothness of the functions $\zeta \to \mathscr{L}_\zeta, v_\zeta, h_\zeta, \beta(\zeta)$.

Given (3.9) the rest of the proof is fairly routine. Choose $k \in \mathscr{P}$ such that $\int_{-\delta}^{\delta} k(s)\, ds > 1 - \delta$ for a suitably small $\delta > 0$, and such that $\int_{-\infty}^{\infty} |s| k(s)\, ds < \infty$. Since k is even,

$$\int u(s - t)(k * N^i_{n,m})(ds) = \int (k * u)(s - t) N^i_{n,m}(ds).$$

Consequently

$$\left| \int u(s - t)(k * N^i_{n,m})(ds) - \int u(s - t) N^i_{n,m}(ds) \right| \leqslant \int |u(s - t) - (k * u)(s - t)|\, N^i_{n,m}(ds)$$

If $\delta > 0$ is sufficiently small (how small depends on u) then $|u(s) - k * u(s)| < \varepsilon$ for all $s \in [a - 1, b + 1]$, where support $u \subset [a, b]$, while for $s \notin [a - 1, b + 1]$, $u(s) = 0$ and

$$k * u(s) \leqslant \left(\int_{a-s}^{\infty} k(y)\, dy \right) \|u\|_\infty, \; s < a,$$

$$k * u(s) \leqslant \left(\int_{s-b}^{\infty} k(y)\, dy \right) \|u\|_\infty, \; s > b.$$

Now $k \in \mathscr{P}$ may be chosen so that $\int_{-\infty}^{\infty} |s| k(s)\, ds$ is arbitrarily small, and hence

$$\sum_{j=1}^{\infty} \int_{j}^{\infty} k(y)\, dy < \varepsilon.$$

Therefore, by (3.9)

$$\left| \int u(s-t)(k * N^i_{n,m})(ds) - \int u(s-t) N^i_{n,m}(ds) \right|$$

$$\leqslant \varepsilon N^i_{n,m}([a-1, b+1])$$

$$+ \sum_{j=1}^{\infty} \|u\|_{\infty} \int_{j}^{\infty} k(y)\, dy N^i_{n,m}([a-j-1, a-j] \cup [b+j, b+j+1])$$

$$\leqslant C(\xi) n^{-(2g+1)/2} e^{-n\gamma(\xi)+z_0 t}(\varepsilon(b-a+2) + 2\varepsilon\|u\|_{\infty}).$$

Since $\varepsilon > 0$ may be made arbitrarily small by a judicious choice of $k \in \mathscr{P}$, (3.8) implies (3.7). (The local uniformity in ξ follows from (3.9).) □

4. Thermodynamic functions for the suspension flow. Before we can use (3.1)–(3.2) to prove Th. 1 we must relate the thermodynamic functions β, γ for the shift (Σ_A, σ) to corresponding functions for the suspension flow (Σ_A^r, σ^r). For $\Psi: \Sigma_A^r \to \mathbb{R}$ bounded and Borel measurable define

$$\psi(x) = \int_{s=0}^{r(x)} \Psi(x, s)\, ds$$

(thus, upper case letters denote functions on Σ_A^r, lower case letters the corresponding functions on Σ_A). Let $\mathscr{F}_\rho^r = \{\Psi: \psi \in \mathscr{F}_\rho\}$. For real-valued $\Psi \in \mathscr{F}_\rho^r$ define $P(\Psi)$ to be the unique real number such that

$$\lambda_{\psi - P(\Psi)r} = 1;$$

since $\log \lambda_{\psi - pr}$ is a strictly decreasing, continuous function of p that converges to $\mp\infty$ as $p \to \pm\infty$ (cf. (2.2)), $P(\Psi)$ is well defined.

Let $\mathscr{I}(\mathscr{I}^r)$ denote the set of invariant probability measures for the shift (suspension flow). There is a 1-to-1 correspondence between $\mathscr{I}$ and $\mathscr{I}^r$ given by

$$\mu \leftrightarrow \bar{\mu} \text{ iff } \int \Psi\, d\bar{\mu} = \int \psi\, d\mu \Big/ \int r\, d\mu \ \forall \Psi \in \mathscr{F}_\rho^r.$$

If $\mu \leftrightarrow \bar{\mu}$ then

$$H(\bar{\mu}, \sigma^r) = H(\mu, \sigma) \Big/ \int r \, d\mu. \tag{4.1}$$

(here H denotes entropy; cf. [1]). As before there is a Gibbs variational principle relating the pressure and entropy functions P and H: if $\Psi \in \mathscr{F}_\rho^r$ is real-valued, then for any $\bar{\mu} \in \mathscr{I}^r$

$$P(\Psi) \geqslant H(\bar{\mu}, \sigma^r) + \int \Psi \, d\bar{\mu} \tag{4.2}$$

with equality iff $\bar{\mu} = \bar{\mu}_\Psi$, where $\bar{\mu}_\Psi \leftrightarrow \mu_{\psi - P(\Psi)r}$ (see [6] or [7]).

Consider now the functions $\Phi_1, \ldots, \Phi_{2g}$ on Σ_A^r constructed in sec. 1. Define

$$B(z) = P\left(\sum_1^{2g} z_i \Phi_i\right), \qquad z \in \mathbb{R}^{2g}, \tag{4.3}$$

$$\Gamma(\xi) = \sup_{z \in \mathbb{R}^{2g}} (\langle \xi, z \rangle - B(z)), \qquad \xi \in \mathbb{R}^{2g}; \tag{4.4}$$

observe that B, β satisfy

$$\beta(-B(z), z_1, \ldots, z_{2g}) = 0 \qquad \forall z = (z_1, \ldots, z_{2g})^t \in \mathbb{R}^{2g}. \tag{4.5}$$

(NOTE: the pressure function β for the original $(r, \varphi_1, \ldots, \varphi_{2g})$ in sec. 1 is identical to that for the cohomologue $(r^*, \varphi_1^*, \ldots, \varphi_{2g}^*)$ introduced at the end of sec. 1).

PROPOSITION 5. *Fix $z = (z_1, \ldots, z_{2g})^t \in \mathbb{R}^{2g}$; let $\xi = \nabla B(z)$, $\xi^* = (1, \xi_1, \ldots, \xi_{2g})^t$, and $t_\xi = \int r \, d\mu$ where $\mu = \mu_{\Sigma z_i \varphi_i - B(z)r}$ is the Gibbs measure on Σ_A for the function $\sum_{i=1}^{2g} z_i \varphi_i - B(z)r$. Let $\bar{\mu} = \bar{\mu}_{\Sigma z_i \Phi_i}$ be the invariant measure for the suspension flow such that $\mu \leftrightarrow \bar{\mu}$. Then*

$$\xi = \nabla B(z) = \left(\int \Phi_1 \, d\bar{\mu}, \ldots, \int \Phi_{2g} \, d\bar{\mu} \right)^t; \tag{4.6}$$

$$\nabla^2 B(z) \text{ is strictly positive definite}; \tag{4.7}$$

$$\det(\nabla^2 B(z)) = t_\xi^{-2g} \det(\nabla^2 \beta(-B(z), z_1, \ldots, z_{2g})) \langle \xi^*, \nabla^2 \gamma(t_\xi \xi^*) \xi^* \rangle. \tag{4.8}$$

The proof uses (4.5) and properties of the functions β, γ obtained in sec. 2. We shall defer it until the end of this section.

Observe that (4.7) implies that B is strictly convex on $\mathbb{R}^{2g}$ and that ∇B is a diffeomorphism of $\mathbb{R}^{2g}$ onto an open subset $\bar{\Omega}$ of $\mathbb{R}^{2g}$. Consequently, Γ, the Legendre

transform of B, is finite and convex on $\bar{\Omega}$. For $\xi \in \bar{\Omega}$, $\xi = \nabla B(z)$, the sup in (4.4) is attained uniquely at z. Therefore, $\nabla\Gamma$ is smooth on $\bar{\Omega}$ and

$$\nabla\Gamma \circ \nabla B = \text{identity on } \mathbb{R}^{2g},$$

$$\nabla B \circ \nabla\Gamma = \text{identity on } \bar{\Omega},$$

$$\nabla^2\Gamma(\xi) = \nabla^2 B(z)^{-1} \text{ if } \xi = \nabla B(z).$$

The Gibbs variational principle implies that

$$\Gamma(\xi) = -H(\bar{\mu}_{\Sigma z_i \Phi_i}, \sigma^r) \text{ if } \nabla B(z) = \xi, \tag{4.9}$$

and thus that $-\Gamma(\xi)$ is the maximum entropy of any invariant measure $\bar{\mu}$ satisfying $\int \Phi \, d\bar{\mu} = \xi$.

PROPOSITION 6. *Fix $z \in \mathbb{R}^{2g}$; let $\xi = \nabla B(z)$, $\xi^* = (1, \xi_1, \xi_2, \ldots, \xi_{2g})^t$, and $t_\xi = \int r \, d\mu$ where $\mu = \mu_{\Sigma z_i \varphi_i - B(z)r}$ is the Gibbs measure for $\Sigma z_i \varphi_i - B(z)r$. Then*

$$\Gamma(\xi) = \gamma(t_\xi \xi^*)/t_\xi = \inf_{t>0} \gamma(t\xi^*)/t; \tag{4.10}$$

$$(d^2/dt^2)(\gamma(t\xi^*)/t)_{t=t_\xi} = t_\xi^{-1} \langle \xi^*, \nabla^2\gamma(t_\xi \xi^*)\xi^* \rangle; \tag{4.11}$$

$$\det \nabla^2\Gamma(\xi) = t_\xi^{2g} \langle \xi^*, \nabla^2\gamma(t_\xi \xi^*)\xi^* \rangle^{-1} \det \nabla^2\gamma(t_\xi \xi^*). \tag{4.12}$$

$$\nabla^2\Gamma(\xi) \text{ is strictly positive definite.} \tag{4.13}$$

Proof. Notice first that $t_\xi \xi^* = (\int r \, d\mu, \int \varphi_1 \, d\mu, \ldots, \int \varphi_{2g} \, d\mu) = \nabla\beta(-B(z), z_1, \ldots, z_{2g})$, so $t_\xi \xi^* \in \Omega$. Recall (sec. 2) that γ is C^∞ in Ω. By the chain rule

$$(d/dt)(\gamma(t\xi^*)/t) = -t^{-2}\gamma(t\xi^*) + t^{-1}\langle \nabla\gamma(t\xi^*), \xi^* \rangle$$

for t near t_ξ; evaluated at $t = t_\xi$ this derivative is 0 because $-\gamma(t_\xi \xi^*) + \langle \nabla\gamma(t_\xi \xi^*), t_\xi \xi^* \rangle = \beta(\nabla\gamma(t_\xi \xi^*)) = \beta(-B(z), z_1, \ldots, z_{2g}) = 0$. Differentiating a second time and using the fact that the first derivative is 0 at $t = t_\xi$ gives (4.11). Since R.H.S. (4.11) > 0, it follows that $t \to \gamma(t\xi^*)/t$ has a local minimum at $t = t_\xi$. But since γ is convex, $s \to s\gamma(\xi^*/s)$ is also convex for $s > 0$, so $t \to \gamma(t\xi^*)/t$ has a global minimum at $t = t_\xi$. The fact that $\gamma(t_\xi \xi^*)/t_\xi = \Gamma(\xi)$ now follows from the Gibbs variational principles for the shift and the flow, together with (4.1). This proves (4.10). Finally, (4.12)–(4.13) follow immediately from (4.7)–(4.8) since $\nabla^2\Gamma(\xi) = \nabla^2 B(z)^{-1}$ and $\nabla^2\gamma(t_\xi \xi^*) = \nabla^2\beta(-B(z), z_1, \ldots, z_{2g})^{-1}$. □

Note that the preceding results have been established only for $\xi \in \bar{\Omega}$, not for all $\xi \in \mathbb{R}^{2g}$. Unfortunately, it does not seem possible to describe $\bar{\Omega}$ completely; however,

PROPOSITION 7. *$\bar{\Omega}$ contains a neighborhood of the origin; also $-\Gamma(0)$ = topological entropy of σ^r.*

Proof. Since $\bar{\Omega}$ is open, it suffices to show that $0 \in \bar{\Omega}$. Consider the maximum entropy invariant measure ν for the geodesic flow on the unit tangent bundle SM; it follows from Bowen's symbolic dynamics (specifically, (1.1)–(1.3)) that ν is unique. Let I: $SM \to SM$ be the map that reverses directions, i.e., $I(m, \theta) = (m, -\theta)$, and let $\tilde{\nu} = \nu \circ I$; clearly, $\tilde{\nu}$ is an invariant measure for the geodesic flow with the same entropy as ν. Therefore, $\nu = \tilde{\nu}$. It follows that $\int W_i \, d\nu = 0$ for $i = 1, 2, \ldots, 2g$, because $W_i \circ I = -W_i$, so $\int W_i \, d\nu = -\int W_i \, d\tilde{\nu}$.

The measure $\bar{\mu} = \nu \circ \pi$ on Σ_A^r is the maximum entropy invariant measure for the suspension flow. The result of the previous paragraph implies that $\int \Phi_i \, d\bar{\mu} = 0$ for $i = 1, \ldots, 2g$. Now $H(\bar{\mu}, \sigma^r)$ = topological entropy of σ^r; but it is known ([11], Prop. 2) that the topological entropy of σ^r equals $P(0)$. It follows from the variational principle that $\bar{\mu} = \bar{\mu}_0$. Prop. 5 now implies that $\nabla B(0) = (\int \Phi_1 \, d\bar{\mu}, \ldots, \int \Phi_{2g} \, d\bar{\mu}) = 0$, so $0 \in \bar{\Omega}$. □

Proof of Prop. 5. Taking the partial derivative with respect to z_i in (4.5) gives

$$\partial_i \beta(-B(z), z) = \partial_0 \beta(-B(z), z)(\partial/\partial z_i) B(z)$$

$$\Rightarrow \partial B/\partial z_i = \partial_i \beta(-B(z), z)/\partial_0 \beta(-B(z), z)$$

$$= \int \varphi_i \, d\mu \Big/ \int r \, d\mu = \int \Phi_i \, d\bar{\mu},$$

proving (4.6) (cf. (2.2)). Taking another partial derivative, this time with respect to z_j, gives

$$\partial_0 \beta(-B(z), z)(\partial^2 B/\partial z_i \partial z_j) = \partial_{00} \beta(-B(z), z)(\partial B/\partial z_i)(\partial B/\partial z_j) + \partial_{ij} \beta(-B(z), z)$$

$$- \partial_{0i} \beta(-B(z), z)(\partial B/\partial z_j) - \partial_{0j} \beta(-B(z), z)(\partial B/\partial z_i)$$

This may be rewritten as

$$A = C^t D C,$$

where

$$A = \partial_0 \beta(-B(z), z) \nabla^2 B(z),$$

$$D = \nabla^2 \beta(-B(z), z),$$

$$C = (c_{ij}),\ i = 0, 1, \ldots, 2g,\ j = 1, 2, \ldots, 2g,$$

$$c_{ij} = \begin{cases} -\xi_j, & i = 0 \\ \delta_{ij}, & i = 1, \ldots, 2g. \end{cases}$$

(Here δ_{ij} is the Kronecker delta.) Since D is strictly positive definite (2.3), so is A; this proves (4.7).

Let $v = D^{-1}\xi^* \in \mathbb{R}^{2g+1}$; it is easily verified that $C^tDv = 0 \in \mathbb{R}^{2g}$. Consequently,

$$\begin{pmatrix} v^t \\ C^t \end{pmatrix} D(v|C) = \left(\begin{array}{c|c} (\xi^*)^t D^{-1}\xi^* & 0^t \\ 0 & A \end{array}\right)$$

$$\Rightarrow (\det D)(\det(v|C))^2 = ((\xi^*)^t D^{-1}\xi^*) \det A.$$

Now

$$(v|C) = \begin{bmatrix} v_0 & -\xi_1 & -\xi_2 & -\xi_3 & \cdots \\ v_1 & 1 & 0 & 0 & \cdots \\ v_2 & 0 & 1 & 0 & \cdots \\ v_3 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix};$$

elementary column operations show that its determinant is $\langle v, \xi^* \rangle = \langle \xi^*, D^{-1}\xi^* \rangle$. Thus

$$\det(\nabla^2\beta(-B(z), z))\langle \xi^*, \nabla^2\beta(-B(z), z)^{-1}\xi^* \rangle = \partial_0\beta(-B(z), z)^{2g} \det(\nabla^2 B(z))$$

which proves (4.8) since $\partial_0\beta(-B(z), z) = \int r \, d\mu = t_\xi$ and $\nabla^2\beta(-B(z), z)^{-1} = \nabla^2\gamma(\xi^*)$. □

5. The final tally

PROPOSITION 8. *Let $t^{-1}(m_1, \ldots, m_{2g}) = (\xi_1, \ldots, \xi_{2g}) = \xi$. Uniformly for ξ in some neighborhood of the origin, as $t \to \infty$*

$$\sum_{n=1}^{\infty} n^{-1}Q_n(t, m; \tau) \sim e^{-t\Gamma(\xi)}t^{-g-1}(2\pi)^{-g}(\det \nabla^2\Gamma(\xi))^{1/2}C_\xi(\tau), \text{ where} \tag{5.1}$$

$$C_\xi(\tau) = \int_0^\tau \exp\{-(\Gamma(\xi) - \langle \nabla\Gamma(\xi), \xi \rangle)s\} \, ds. \tag{5.2}$$

Proof. By Prop. 7 there is a neighborhood of the origin contained in $\bar{\Omega}$. Assume that ξ is in this neighborhood, and let z, ξ^*, t_ξ, $\mu = \mu_{\Sigma z_i\varphi_i - B(z)r}$ be as in Prop. 5. We will show later, using Prop. 4, that for any $\delta > 0$, as $t \to \infty$

$$\sum_{n=1}^{\infty} n^{-1}Q_n(t, m; \tau) \sim \sum_{|n - t/t_\xi| \leqslant \delta t} n^{-1}Q_n(t, m; \tau) \tag{5.3}$$

uniformly in ξ near 0.

Prop. 3 gives an asymptotic formula for Q_n; substituting this in (5.3) gives

$$(5.4)\qquad \sum_{|n-t/t_\xi|\leqslant \delta t} n^{-1}Q_n(t, m; \tau) \sim \sum_{|n-t/t_\xi|\leqslant \delta t} \exp\{-t(n/t)\gamma((t/n)\xi^*)\}n^{-g-3/2}(2\pi)^{-g-1/2}$$
$$\cdot(\det \nabla^2\gamma((t/n)\xi^*))^{1/2}\int_0^\tau e^{-z_n s}\,ds$$

where $z_n = \partial_0\gamma((t/n)\xi^*)$. By Prop. 6, $s \to \gamma(s\xi^*)/s$ has its minimum at $s = t_\xi$, where it has a positive second derivative. Consequently, Taylor's theorem implies that

$$\exp\{-t(n/t)\gamma((t/n)\xi^*)\} \sim \exp\{-t\Gamma(\xi)\}\exp\{-tb_\xi((t/n) - t_\xi)^2/2\}$$

uniformly for n such that $|n - t/t_\xi| \leqslant t^{1/2}\log t$, where

$$b_\xi = (d^2/ds^2)(\gamma(s\xi^*)/s)$$
$$= t_\xi^{-1}\langle \xi^*, \nabla^2\gamma(t_\xi\xi^*)\xi^*\rangle,$$

and

$$\exp\{-t(n/t)\gamma((t/n)\xi^*)\} \leqslant \exp\{-t\Gamma(\xi)\}\exp\{-tb_\xi((t/n) - t_\xi)^2/4\}$$

for all n such that $|n - t/t_\xi| \leqslant \delta t$, provided $\delta > 0$ is sufficiently small. Thus the major contribution to the sum in (5.4) comes from the terms for which $|n - t/t_\xi| \leqslant t^{1/2}\log t$, and the sum, suitably renormalized, is a Riemann sum for $\int_{-\infty}^{\infty} e^{-u^2/2}\,du$ (use $u = t^{1/2}b_\xi^{1/2}((t/n) - t_\xi)$). It follows that

$$\sum_{|n-t/t_\xi|\leqslant \delta t} n^{-1}Q_n(t, m; \tau)$$
$$\sim e^{-t\Gamma(\xi)}t^{-g-1}(2\pi)^{-g}\langle \xi^*, \nabla^2\gamma(t_\xi\xi^*)\xi^*\rangle^{-1/2}(\det \nabla^2\gamma(t_\xi\xi^*))^{1/2}\cdot t_\xi^g\int_0^\tau e^{-z_0 s}\,ds$$

where $z_0 = \partial_0\gamma(t_\xi\xi^*) = -B(\nabla\Gamma(\xi)) = -(\Gamma(\xi) - \langle\nabla\Gamma(\xi), \xi\rangle)$. This, together with (4.12), proves (5.1), modulo the proof of (5.3).

Recall again that $s \to \gamma(s\xi^*)/s$ has its minimum at $s = t_\xi$, where the second derivative is positive. Now $t_\xi\xi^* \in \Omega$ (see the proof of Prop. 6) so $s\xi^* \in \Omega$ for s near t_ξ; for such s

$$(d/ds)(\gamma(s\xi^*)/s) = s^{-2}(-\gamma(s\xi^*) + \langle\nabla\gamma(s\xi^*), s\xi^*\rangle)$$
$$= s^{-2}\beta(\nabla\gamma(s\xi^*)),$$

hence

$$\beta(\nabla\gamma(s\xi^*)) \gtrless 0 \qquad \text{for } s \gtrless t_\xi.$$

Let $n_0 = [[t(t_\xi^{-1} - \delta)]]$ and $n_1 = [[t(t_\xi^{-1} + \delta)]]$ (here $[[\cdot]]$ denotes greatest integer); let $s_i = t/n_i$, $i = 0, 1$, and let $z^i = \nabla\gamma(s_i\xi^*)$, $i = 0, 1$. Since $s_0 > t_\xi$ and $s_1 < t_\xi$ we have $\beta(z^0) > 0$ and $\beta(z^1) < 0$, by the preceding paragraph. Thus, if $n \leqslant n_0$ then Prop. 4 (with $\zeta = (t, m_1, \ldots, m_{2g}) = t\xi^*$) implies

$$Q_n(t, m; \tau) \leqslant K_{z^0} \exp\{n\beta(z^0) - \langle z^0, \zeta\rangle + |z_0^0|\tau\}$$

$$\Rightarrow \sum_{n=1}^{n_0} Q_n(t, m; \tau) \leqslant K_{z^0} \exp\{n_0\beta(z^0) - \langle z^0, \zeta\rangle + |z_0^0|\tau\}/(1 - e^{-\beta(z^0)})$$

$$= K_{z^0} \exp\{(t/s_0)(\beta(z^0) - \langle z^0, s_0\xi^*\rangle) + |z_0^0|\tau\}/(1 - e^{-\beta(z^0)})$$

$$= K_{z^0} \exp\{-(t/s_0)\gamma(s_0\xi^*) + |z_0^0|\tau\}/(1 - e^{-\beta(z^0)}).$$

Since $-\gamma(s_0\xi^*)/s_0 < -\gamma(t_\xi\xi^*)/t_\xi = -\Gamma(\xi)$, this proves that

$$\sum_{n=1}^{n_0} Q_n(t, m; \tau) = o\left(\sum_{n=n_0}^{n_1} Q_n(t, m; \tau)\right).$$

Similarly, if $n \geqslant n_1$ then Prop. 4 implies

$$Q_n(t, m; \tau) \leqslant K_{z^1} \exp\{n\beta(z^1) - \langle z^1, \zeta\rangle + |z_0^1|\tau\}$$

$$\Rightarrow \sum_{n=1}^{\infty} Q_n(t, m; \tau) \leqslant K_{z^1} \exp\{n_1\beta(z^1) - \langle z^1, \zeta\rangle + |z_0^1|\tau\}/(1 - e^{\beta(z^1)})$$

$$= K_{z^1} \exp\{(t/s_1)(\beta(z^1) - \langle z^1, s_1\xi^*\rangle) + |z_0^1|\tau\}/(1 - e^{\beta(z^1)})$$

$$= K_{z^1} \exp\{-(t/s_1)\gamma(s_1\xi^*) + |z_0^1|\tau\}/(1 - e^{\beta(z^1)});$$

again, $-\gamma(s_1\xi^*)/s_1 < -\Gamma(\xi)$, so

$$\sum_{n=n_1}^{\infty} Q_n(t, m; \tau) = o\left(\sum_{n=n_0}^{n_1} Q_n(t, m; \tau)\right).$$

This proves (5.3). Observe that all the estimates hold uniformly for ξ locally, since the thermodynamic functions are continuous in ξ. □

Theorem 1 follows easily from Proposition 8. First we argue that (5.1) implies

$$\sum_{n=1}^{\infty} n^{-1}\tilde{Q}_n(t, m; \tau) \sim R.H.S.\ (5.1), \tag{5.5}$$

where

$$\tilde{Q}_n(t, m; \tau) = \sum_{x \in \mathscr{P}_n} 1\{0 \leqslant S_n r(x) - t \leqslant \tau; S_n \varphi_i(x) = m_i \ \forall i = 1, \ldots, 2g\}.$$

Recall (cf. (1.5)) that $\mathscr{P}_n$ is the set of periodic sequences x with *least* period n; the difference between Q_n and $\tilde{Q}_n$ is that Q_n counts *all* x such that $\sigma^n x = x$. Hence

$$Q_n(t, m; \tau) - \tilde{Q}_n(t, m; \tau) \leqslant \sum_{d|n, d>1} Q_{n/d}(t/d, m/d; \tau)$$

$$\Rightarrow \sum_{n=1}^{\infty} n^{-1}(Q_n(t, m; \tau) - \tilde{Q}_n(t, m; \tau)) \leqslant \sum_{2 \leqslant d \leqslant Ct} \sum_{n=1}^{\infty} n^{-1} Q_n(t/d, m/d; \tau)$$

(note that $Q_n(t/d, m/d; \tau) = 0$ unless $m/d \in \mathbb{Z}^{2g}$; also, $d \leqslant Ct$ because there is a $\delta > 0$ such that every periodic orbit of the suspension flow has length $\geqslant \delta$, so $t/d + \tau \geqslant \delta$). But (5.1) implies that the last (double) sum is $\leqslant C' t \exp\{-t\Gamma(\xi)/2\}$, which is of smaller exponential order of magnitude than R.H.S. (5.1). This proves (5.5).

Consider the quantity $R(t; m_1, \ldots, m_{2g}) = R(t; m)$ defined by (1.5); we may write

$$R(t; m) - R(t - K\tau t; m) = \sum_{k=1}^{[[Kt]]} \left(\sum_{n=1}^{\infty} n^{-1} \tilde{Q}_n(t - k\tau, m; \tau) \right).$$

If $\xi = t^{-1}m$ is sufficiently close to the origin and $K\tau$ is not too large (recall that (5.1) and hence (5.5) may only be valid for ξ near the origin) then each of the terms in the above sum may be estimated by (5.5), yielding

$$(5.6) \quad R(t; m) - R(t - K\tau t; m) \sim \sum_{k=1}^{[[Kt]]} \exp\{-(t - k\tau)\Gamma((t - k\tau)^{-1}m)\}(t - k\tau)^{-g-1} \cdot \det \nabla^2 \Gamma((t - k\tau)^{-1}m)^{1/2} C_{(t-k\tau)^{-1}m}(\tau).$$

Observe that $(d/dt)(t\Gamma(t^{-1}m)) = \Gamma(t^{-1}m) - \langle \nabla\Gamma(t^{-1}m), t^{-1}m \rangle = -B(\nabla\Gamma(t^{-1}m))$; if $t^{-1}m$ is near the origin then $\nabla\Gamma(t^{-1}m)$ is near the origin, because $\nabla B(0) = 0$ (Prop. 7), and thus $-B(\nabla\Gamma(t^{-1}m)) < 0$, because $B(0) > 0$ by (4.5). Hence the terms in the above series are exponentially decreasing, and the major contribution comes from the range $1 \leqslant k \leqslant \sqrt{t/\tau}$. Using (5.2) and the above formula for $(d/dt)(t\Gamma(t^{-1}m))$ we obtain

$$R(t; m) - R(t - K\tau t; m) \sim e^{-t\Gamma(\xi)} t^{-g-1} (2\pi)^{-g} (\det \nabla^2 \Gamma(\xi))^{1/2} \cdot (\langle \nabla\Gamma(\xi), \xi \rangle - \Gamma(\xi))^{-1}.$$

Finally, we argue that if $\xi = t^{-1}m$ is in a sufficiently small neighborhood of the origin and if $K\tau$ is suitably chosen then $R(t - K\tau t; m) = O(\exp\{-t\Gamma(\xi) - t\varepsilon\})$ for some $\varepsilon > 0$. If ξ is near the origin then $-\Gamma(\xi)$ is near the topological entropy h of the flow (Prop. 7). Thus $K\tau$ may be chosen small enough that (5.6) is valid, but large

enough that $(1 - K\tau)h \leqslant -\Gamma(\xi) - \varepsilon$ for some $\varepsilon > 0$. Now $R(t - K\tau t; m) \leqslant$ the total number of periodic orbits of the flow with period $\leqslant (1 - K\tau)t$, which is $\leqslant$ $(\text{const})e^{t(1-K\tau)h}$ by Margulis's theorem [9].

This proves that

$$R(t; m) \sim e^{-t\Gamma(\xi)}t^{-g-1}(2\pi)^{-g}(\det \nabla^2\Gamma(\xi))^{1/2}(\langle \nabla\Gamma(\xi), \xi\rangle - \Gamma(\xi))^{-1}$$

provided $t^{-1}m = \xi$ is in a sufficiently small neighborhood of the origin. All of the above approximations hold uniformly in ξ locally, by the continuity of the thermodynamic functions. This completes the proof of Theorem 1.

Appendix: Proof of Proposition 0. The strategy will be to alter the sequence space Σ_A furnished by the Bowen/Ratner construction so as to obtain a new sequence space $\Sigma_{\bar{A}}$. This new sequence space will be constructed in such a way that all "periodicities" in the functions $\varphi_1^*, \ldots, \varphi_{2g}^*$ are destroyed.

Step 1. *Enlarging the alphabet.* The alphabet for the sequence space Σ_A is $\mathscr{A} = \{1, 2, \ldots, l\}$. Let $k > 1$ be an integer; define $\mathscr{A}_k$ to be the set of all sequences of length k from $\mathscr{A}$ with transitions allowed by A, i.e., $\mathscr{A}_k = \{x_1x_2\ldots x_k\colon x_i \in \mathscr{A}$ and $A(x_i, x_{i+1}) = 1\}$. Define a transition matrix A_k on $\mathscr{A}_k$ by

$$A_k(x_1x_2\ldots x_k, x_1'x_2'\ldots x_k') = \begin{cases} 1 & \text{if } A(x_k, x_k') = 1 \text{ and } x_i' = x_{i+1}\ \forall i = 1, \ldots, k-1; \\ 0 & \text{otherwise.} \end{cases}$$

Let $q_k\colon \mathscr{A}_k \to \mathscr{A}$ be the projection on the first coordinate (i.e., $q_k(x_1x_2\ldots x_k) = x_1$) and let $p_k\colon \Sigma_{A_k} \to \Sigma_A$ be the induced map on sequence space. Clearly, p_k is bijective and commutes with the shift; moreover, for each $\rho \in (0, 1)$ the maps p_k and p_k^{-1} are Lipschitz relative to the metrics d_ρ on Σ_{A_k} and Σ_A. Hence, each of the sequence spaces Σ_{A_k} provides an alternative "symbolic dynamics" for the geodesic flow.

The reason for introducing the spaces Σ_{A_k} is that they provide much enlarged alphabets. In particular, if $\{x^1, x^2, \ldots, x^m\}$ is any finite collection of periodic sequences in Σ_A, then for all sufficiently large k the periodic sequences $p_k^{-1}(x^1), \ldots, p_k^{-1}(x^m)$ in Σ_{A_k} are such that no two share a common symbol from $\mathscr{A}_k$.

In step 2 we will assume that the original sequence space Σ_A has been replaced by Σ_{A_k} for some k; for ease of notation we will drop the subscript k and write the alphabet as $\mathscr{A} = \{1, 2, \ldots, l\}$. In step 3 we will specify k.

Step 2. *Inserting "loops" in sequences.* For each symbol $i \in \mathscr{A}$ we invent new symbols $i_1, i_2, \ldots, i_{M(i)}$ and let $\bar{\mathscr{A}}$ be the alphabet consisting of all the new symbols. Thus, $\bar{\mathscr{A}} = \{1_1, 1_2, \ldots, 1_{M(1)}, 2_1, 2_2, \ldots, l_{M(l)}\}$. Define a transition matrix $\bar{A}$ on $\bar{\mathscr{A}}$ by

$$\bar{A}(i_j, i_{j+1}) = 1, \qquad j = 1, 2, \ldots, M(i) - 1;$$

$$\bar{A}(i_{M(i)}, i_1') = 1 \qquad \text{if } A(i, i') = 1;$$

$$\bar{A}(i_j, i_{j'}') = 0 \qquad \text{otherwise.}$$

For any sequence in Σ_A there is a unique sequence in $\Sigma_{\bar{A}}$ obtained by replacing each symbol $i \in \mathscr{A}$ by the word $i_1 i_2 \dots i_{M(i)}$. Conversely, for any sequence in $\Sigma_{\bar{A}}$ there is a corresponding sequence in Σ_A obtained by deleting all symbols in the sequence except $1_1, 2_1, \dots, l_1$, then removing the subscript 1 on each of the remaining symbols, then applying σ^{-1}. Thus, there is a surjective map $p: \Sigma_{\bar{A}} \to \Sigma_A$; this map is not $1-1$ and does not commute with the shift, unless $M(1) = M(2) = \cdots = M(l) = 1$, but it is continuous. Furthermore, for each $\rho \in (0, 1)$ there exists $\bar{\rho} \in (0, 1)$ such that p is Lipschitz relative to the metrics d_ρ, $d_{\bar{\rho}}$ on Σ_A, $\Sigma_{\bar{A}}$. Most important, p induces a bijection between equivalence classes of periodic sequences in $\Sigma_{\bar{A}}$ and Σ_A (here periodic sequences $x, x' \in \Sigma_A$ or $x, x' \in \Sigma_{\bar{A}}$ are considered equivalent if they are in the same orbit, i.e., if $x' = \sigma^j x$ for some j).

Now consider the functions $r^*, \varphi_1^*, \dots, \varphi_{2g}^*$ on Σ_A; recall that each of these depends only on the forward coordinates $x_0 x_1 x_2 \dots$ of $x \in \Sigma_A$, hence may be considered a function on Σ_A^+. For $x \in \Sigma_{\bar{A}}$ define

$$g(x) = M((p(x))_0),$$

$$\bar{r}^*(x) = r^*(p(x))/g(x),$$

$$\bar{\varphi}_j^*(x) = \begin{cases} \varphi_j^*(p(x)) & \text{if } x_0 \in \{1_1, 2_1, \dots, l_1\}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\bar{r}^* > 0$, $\bar{\varphi}_j^*$ is integer-valued, and $\bar{r}^*, \bar{\varphi}_1^*, \dots, \bar{\varphi}_{2g}^*$ are Lipschitz relative to $d_{\bar{\rho}}$. Let $x \in \Sigma_{\bar{A}}$ be a periodic sequence with minimal period $\bar{n}$, and suppose the minimal period of $p(x)$ is n. Then

$$S_{\bar{n}} \bar{r}^*(x) = S_n r^*(p(x)) \qquad \text{and}$$

$$S_{\bar{n}} \bar{\varphi}_j^*(x) = S_n \varphi_j^*(p(x)), \qquad j = 1, 2, \dots, 2g,$$

so the R.H.S. of (1.5) remains unchanged if we substitute $\bar{r}^*$ for r^*, $\bar{\varphi}_j^*$ for φ_j^*, and $\bar{\mathscr{P}}_n$ for $\mathscr{P}_n$, where $\bar{\mathscr{P}}_n$ is the set of periodic sequences in $\Sigma_{\bar{A}}$ with minimal period n.

It is clear that the new sequence space $\Sigma_{\bar{A}}$ gives an alternative symbolic dynamics for the geodesic flow. The useful feature of this new symbolic dynamics is that the minimal period of the periodic sequence representing a particular closed geodesic can be adjusted by tinkering with the integers $M(1), M(2), \dots, M(l)$.

Step 3. *Removing the periodicities.* We start with the sequence space Σ_A provided by the Bowen/Ratner construction, and let $\varphi_1^*, \varphi_2^*, \dots, \varphi_{2g}^*$ be as in sec. 1. Recall (1.4) that all but finitely many closed geodesics have the property that the preimage (under π) consists of a single periodic orbit of the suspension flow with the same minimal period. Recall also ([2], sec. 11.7, Th. 10) that each homology class $m \in \mathbb{Z}^{2g} \cong H_1 M$ contains a closed geodesic. Consequently, there exist $m^0, m^1, \dots, m^{2g} \in \mathbb{Z}^{2g}$ and periodic sequences $x^0, x^1, \dots, x^{2g} \in \Sigma_A$ with minimal periods $n_0, n_1, \dots, n_{2g}$ such that

(a) $S_{n_i}\varphi^*(x^i) = m^i$ for each $i = 0, 1, \ldots, 2g$; and

(b) the vectors $m^i - m^0$, $i = 1, \ldots, 2g$, are the standard unit vectors in $\mathbb{Z}^{2g}$.

Next, we replace the sequence space Σ_A by Σ_{A_k} as in step 1. Regardless of what $k \geqslant 2$ is used, the sequences $x^0, x^1, \ldots, x^{2g}$ pull back to periodic sequences $\tilde{x}^0, \tilde{x}^1, \ldots, \tilde{x}^{2g}$ in Σ_{A_k} with the same minimal periods $n_0, n_1, \ldots, n_{2g}$. Furthermore, if $\tilde{\varphi}^*$ is the pullback of φ^* to Σ_{A_k} then $S_{n_i}\tilde{\varphi}^*(\tilde{x}^i) = m^i$ for $i = 0, 1, \ldots, 2g$, as before.

Choose k so large that no two of the sequences $\tilde{x}^0, \tilde{x}^1, \ldots, \tilde{x}^{2g}$ share a common symbol. Each period of $\tilde{x}^i$ contains symbols $a_{i1}, a_{i2}, \ldots, a_{iv(i)}$, which occur with frequencies $k_{i1}, k_{i2}, \ldots, k_{iv(i)}$; thus, the (minimal) period of x^i is $k_{i1} + k_{i2} + \cdots + k_{iv(i)}$. Observe that the symbols a_{ij}, $i = 0, 1, \ldots, 2g$ and $j = 1, \ldots, v(i)$, are distinct.

LEMMA. *There exist integers $M(a_{ij}) \geqslant 1$ for $i = 0, 1, \ldots, 2g$ and $j = 1, 2, \ldots, v(i)$ and an integer $\bar{n}$ such that for each $i = 0, 1, \ldots, 2g$*

$$\text{(A1)} \qquad \sum_{j=1}^{v(i)} M(a_{ij})k_{ij} = \bar{n}.$$

The proof is deferred to the end of the appendix. The function $M(\cdot)$ may be extended to the entire alphabet $\mathscr{A}_k$ by setting $M(\alpha) = 1$ for any α not contained in $\{a_{ij}\}$.

Now we use the function $M(\cdot)$ to define the sequence space $\Sigma_{\bar{A}}$ as in step 2 (using the alphabet $\mathscr{A}_k$). The functions $\tilde{r}^*$ and $\tilde{\varphi}_j^*$ pull back to functions $\bar{r}^*$ and $\bar{\varphi}_j^*$ on $\Sigma_{\bar{A}}$ as explained in Step 2, and the R.H.S. of (1.5) remains unchanged when $\bar{r}^*$, $\bar{\varphi}_j^*$, $\bar{\mathscr{P}}_n$ are substituted for r, φ_j, and $\mathscr{P}$. We will show that the vector-valued function $\bar{\varphi}^* = (\bar{\varphi}_1^*, \ldots, \bar{\varphi}_{2g}^*)$ on $\Sigma_{\bar{A}}$ is not cohomologous to any function ψ valued in a coset of a proper subgroup of $\mathbb{Z}^{2g}$; this will complete the proof of Prop. 0.

Consider the periodic sequences $\tilde{x}^0, \tilde{x}^1, \ldots, \tilde{x}^{2g}$ in Σ_{A_k}. There are periodic sequences $\bar{x}^0, \bar{x}^1, \ldots, \bar{x}^{2g}$ in $\Sigma_{\bar{A}}$ that project to $\tilde{x}^0, \tilde{x}^1, \ldots, \tilde{x}^{2g}$; by (A1) the minimal period of each $\bar{x}^i$ is $\bar{n}$. By construction, $S_{\bar{n}}\bar{\varphi}^*(\bar{x}^i) = S_{n_i}\tilde{\varphi}^*(\tilde{x}^i) = m^i$ for each $i = 0, 1, \ldots, 2g$. Now suppose that $\bar{\varphi}^*$ is cohomologous to a function ψ valued in $h + G$, where G is a subgroup of $\mathbb{Z}^{2g}$. Then $S_{\bar{n}}\bar{\varphi}^*(\bar{x}^i) = S_{\bar{n}}\psi(\bar{x}^i) \in \bar{n}h + G$ for $i = 0, 1, \ldots, 2g$. But $S_{\bar{n}}\bar{\varphi}^*(\bar{x}^i) = m^i$, and $(m^1 - m^0), (m^2 - m^0), \ldots, (m^{2g} - m^0)$ are the standard unit vectors in $\mathbb{Z}^{2g}$; therefore $G = \mathbb{Z}^{2g}$.

Proof of the Lemma. Let $k_1, k_2, \ldots, k_r$ be any finite collection of *positive* integers, and define

$$J = \left\{ \sum_{i=1}^{r} n_i k_i : n_i \in \mathbb{Z} \right\},$$

$$J^+ = \left\{ \sum_{i=1}^{r} n_i k_i : n_i \in \mathbb{Z} \text{ and } n_i \geqslant 1 \right\}.$$

The set J is an ideal, so $J = \{nd : n \in \mathbb{Z}\}$ for some $d \geqslant 1$. We will show that $J \backslash J^+$ is bounded above, i.e., that $nd \in J^+$ for all n sufficiently large.

Without loss of generality we may assume that $d = 1$ (if not, replace $k_1, \ldots, k_r$ by $k_1/d, \ldots, k_r/d$). Choose $s_1, \ldots, s_r \in \mathbb{Z}$ so that $\Sigma s_i k_i = 1$, and set $k = \Sigma k_i \geqslant 1$. Now every $n \in \mathbb{Z}$ may be written uniquely as $n = kx + y$ where $0 \leqslant y < k$, so

$$n = \sum_{i=1}^{r} (x + ys_i)k_i.$$

If n is sufficiently large then $(x + ys_i) \geqslant 1$ for each $i = 1, \ldots, r$, so $n \in J^+$.

Now define

$$J_i = \left\{ \sum_{j=1}^{\nu(i)} s_j k_{ij} : s_j \in \mathbb{Z} \right\},$$

$$J_i^+ = \left\{ \sum_{j=1}^{\nu(i)} s_j k_{ij} : s_j \in \mathbb{Z} \text{ and } s_j \geqslant 1 \right\}.$$

For each $i = 0, 1, \ldots, 2g$, $J_i \backslash J_i^+$ is bounded above. Furthermore, since each J_i is an ideal, so is $\bigcap J_i$; consequently, $\bigcap J_i = \{nd : n \in \mathbb{Z}\}$ for some $d \geqslant 1$, and thus $\bigcap J_i$ is *not* bounded above. It follows that

$$\bigcap_{i=0}^{2g} J_i^+ \neq \varnothing. \qquad \square$$

References

1. L. M. Abramov, *On the entropy of flows*, Dokl. Akad. Nauk. SSSR **128** (5) (1959), 873–876.
2. R. L. Bishop and R. J. Crittenden, *Geometry of Manifolds*, Academic Press, New York, 1964.
3. R. Bowen, *Symbolic dynamics for hyperbolic flows*, Amer. J. Math. **95** (1973), 429–450.
4. ———, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Springer Lecture Notes in Math. **470** (1975).
5. R. Bowen and D. Ruelle, *The ergodic theory of Axiom A flows*, Inv. Math. **29** (1975), 181–202.
6. A. Katsuda and T. Sunada, *Homology of closed geodesics in a negatively curved manifold*, Amer. J. Math. **110** (1988), 145–156.
7. S. Lalley, *Distribution of periodic orbits of symbolic and Axiom A flows*, Adv. Appl. Math. **8** (1987), 154–193.
8. ———, *Renewal theorems in symbolic dynamics, with applications to geodesic flows, noneuclidean tessellations, and their fractal limits*, preprint (1987).
9. G. Margulis, *Applications of ergodic theory to the investigation of manifolds of negative curvature*, Functional Anal. Appl. **3** (1969), 335–336.
10. W. Parry, *Bowen's equidistribution theory and the Dirichlet density theorem*, Ergodic Th. Dynamical Systems **4** (1984), 135–146.
11. W. Parry and M. Pollicott, *An analogue of the prime number theorem for closed orbits of Axiom A flows*, Ann. Math. **118** (1983), 573–591.
12. R. Phillips and P. Sarnak, *Geodesics in homology classes*, Duke Math. J. **55** (1988), 287–297.
13. M. Pollicott, *On the rate of mixing of Axiom A flows*, Inv. Math. **81** (1985), 413–426.
14. ———, *Meromorphic extensions of generalised zeta functions*, Inv. Math. **85** (1986), 147–164.
15. ———, *A complex Ruelle-Perron-Frobenius theorem and two counterexamples*, Ergodic Theory Dynamical Systems **4** (1984), 135–146.

16. ———, to appear in J. Diff. Geom.
17. M. RATNER, *Markov partitions for Anosov flows on n-dimensional manifolds*, Israel J. Math. **15** (1973), 92–114.
18. D. RUELLE, *Statistical mechanics of a one-dimensional lattice gas*, Comm. Math. Phys. **9** (1968), 267–278.
19. ———, *Thermodynamic Formalism*, Addison-Wesley, Reading, Mass, 1978.

DEPARTMENT OF STATISTICS, MATHEMATICAL SCIENCES BUILDING, PURDUE UNIVERSITY, WEST LAFAYETTE, INDIANA 47907