EXPONENTIAL ERROR TERMS FOR GROWTH FUNCTIONS ON NEGATIVELY CURVED SURFACES

Mark Pollicott and Richard Sharp

University of Manchester

Abstract. In this paper we consider two counting problems associated with compact negatively curved surfaces and improve classical asymptotic estimates due to Margulis. In the first, we show that the number of closed geodesics of length at most T has an exponential error term. In the second we show that the number of geodesic arcs (between two fixed points x and y) of length at most T has an exponential error term. The proof is based on a detailed study of the zeta function and Poincaré series and benefits from recent work of Dolgopiat.

0. Introduction

Let V be a compact surface with negative curvature and let $\tilde{V}$ be its universal cover (with the lifted metric d). The fundamental group $\pi_1(V)$ can be realized as a group of isometries Γ acting on $\tilde{V}$ with $\tilde{V}/\Gamma = V$. Given a point $x \in \tilde{V}$, we can introduce the orbital counting function $N(T)$ defined by

$$N(T) = \mathrm{Card}\{g \in \Gamma : d(x, gx) \leq T\}.$$

The compact surface V has a countable infinity of closed geodesics γ (of length $l(\gamma)$) and we can also introduce the closed geodesic counting function

$$\pi(T) = \mathrm{Card}\{\gamma : l(\gamma) \leq T\}.$$

For the special case of surfaces of *constant* negative curvature there are results due to Huber [7], [8] giving both an asymptotic formula and error terms for $N(T)$ and $\pi(T)$, as T tends to infinity. In the more general case of variable negative curvature Margulis again established an asymptotic formula for $N(T)$ and $\pi(T)$ but without error terms [10]. While in the constant curvature case one can make use of harmonic analysis and the Selberg Trace Formula, in the more general setting it is necessary to resort to an approach based on the dynamics of the geodesic flow. In this paper, we shall improve Margulis's original estimates to include the appropriate error terms.

We can associate to V the geodesic flow $\phi_t : SV \to SV$, where SV is the unit tangent bundle of V. We shall let $h > 0$ denote the *topological entropy* of ϕ . Then h is also equal to the exponential growth rate of the area of balls in $\tilde{V}$, i.e.,

$$h = \lim_{T \to +\infty} \frac{1}{T} \log \mathrm{Area}\{y \in \tilde{V} : d(x, y) \leq T\}.$$

Notation. We write $f(T) \sim g(T)$ if $f(T)/g(T) \to 1$ as $T \to +\infty$. We write $f(T) = g(T) + O(h(T))$ if there exists $C > 0$ such that $|f(T) - g(T)| \leq Ch(T)$.

The results of Margulis [10] show that $N(T)$ and $\pi(T)$ have the following first order asymptotic expansions, as $T \to +\infty$:

$$N(T) \sim Ce^{hT} \quad \text{ and } \pi(T) \sim \mathrm{li}(e^{hT}),$$

where $C > 0$ is a constant and where $\mathrm{li}(y)$ is the logarithmic integral

$$\mathrm{li}(y) = \int_2^y \frac{1}{\log u} du \sim \frac{y}{\log y}, \text{ as } y \to +\infty.$$

In this paper we show that the error terms take the form given in the following two theorems.

Theorem 1. *There exists* $0 < c < h$ *such that* $\pi(T) = \mathrm{li}(e^{hT}) + O\left(e^{cT}\right)$.

Theorem 2. *There exists* $0 < c < h$ *and* $C > 0$ *such that* $N(T) = Ce^{hT} + O(e^{cT})$.

We now give a brief outline of the contents of the paper. In section 1, we discuss symbolic coding of the geodesic flow from two viewpoints. The first is a geometric coding best suited to studying $N(T)$ and the second is a dynamical coding better suited to studying $\pi(T)$. In section 2, we give a brief account of Dolgopiat's recent work on transfer operators. In section 3 we use estimates on transfer operators obtain estimates for a certain zeta function. In section 4 we complete the proof of Theorem 1.

In section 5 we introduce the Poincaré series associated to orbit counting and a modified transfer operator. In section 6 prove the necessary results on the associated Poincaré series to complete the proof of Theorem 2.

1. Symbolic Coding for the Geodesic Flow.

In this section we shall describe two methods for coding the geodesic flow and the discrete group Γ. The first approach is the geometric coding obtained by Bowen and Series of Γ and its limit set. The second is the Bowen-Ratner coding for the geodesic flow.

1.1 The geometric coding. Usually the Bowen-Series coding is only applied to geodesic flows on surfaces of constant negative curvature. However, it is an easy observation that it can be adapted to study geodesics flows on surfaces of variable curvature. This is easily seen because every geodesic flow with respect to a metric of variable negative curvature is flow equivalent (i.e. conjugate up to a time change) to one of constant negative curvature [6].

Suppose that V has genus $g \geq 2$ then $\Gamma \cong \pi_1(V)$ has the one relator presentation

$$< a_1, \ldots, a_g, b_1, \ldots, b_g : \prod_{i=1}^{g} [a_i, b_i] = 1 > .$$

We can associate to this presentation a finite directed graph $\mathcal{G}$ with edge set $\mathcal{E}$ and vertex set $\mathcal{V}$ together with:

(1) a distinguished vertex $*$ with no edges ending at $*$;
(2) a labeling of the edges $\lambda : \mathcal{E} \to \Gamma_0 = \{a_1, \ldots, a_g, b_1, \ldots, b_g\}$,

such that there is a bijection between

(i) the set of finite paths starting at $*$; and

(ii) elements of Γ,

given by multiplying the labels on each edge. Furthermore, $\mathcal{G}$ has the important additional property that the map from closed loops to conjugacy classes in Γ obtained by considering the product of the labels around the loop is a bijection [20, 21].

For technical reasons it will prove useful to consider an augmented graph $\mathcal{G}'$. This is obtained by adding an extra vertex 0 and edges $(i, 0)$, for all $i \in \mathcal{V} - \{*\}$. We extend the map λ to the new edges by $\lambda(i, 0) = e$, the identity in Γ.

Let A denote the incidence matrix of $\mathcal{G}'$. Let B denote the submatrix of A obtained by deleting the rows and columns indexed by $*$ and 0.

We can associate one-sided subshifts of finite type $\sigma : X_A^+ \to X_A^+$ and $\sigma : X_B^+ \to X_B^+$ where

$$X_A^+ = \{x = (x_n)_{n=0}^\infty \in \prod_{n=0}^\infty (\mathcal{V} \cup \{0\}) \,:\, A(x_n, x_{n+1}) = 1, \forall n \geq 0\};$$

$$X_B^+ = \{x = (x_n)_{n=0}^\infty \in \prod_{n=0}^\infty (\mathcal{V} - \{*\}) \,:\, B(x_n, x_{n+1}) = 1, \forall n \geq 0\}$$

and $\sigma(x_n)_{n=0}^\infty = (x_{n+1})_{n=0}^\infty$. We can define metrics on X_A^+ and X_B^+ by

$$d((x_n)_{n=0}^\infty, (y_n)_{n=0}^\infty) = \sum_{n=0}^\infty \frac{1 - \delta(x_n, y_n)}{2^n},$$

where $\delta(i, j) = 1$ if $i = j$ and 0 otherwise. With respect to these metrics the spaces X_A^+ and X_B^+ are compact zero dimensional spaces and σ is a local homeomorphism.

Since we are dealing with surfaces the matrix B is aperiodic, or equivalently the subshift $\sigma : X_B^+ \to X_B^+$ is topologically mixing.

1.2 The dynamical coding. We now turn to an alternative approach studied by Sinai, Bowen, Ratner and others. In this case we obtain a symbolic description of the geodesic flow via certain families of local cross sections.

Let SV denote the unit tangent bundle for the surface V. Let $\phi_t : SV \to SV$ denote the geodesic flow on unit tangent vectors (i.e., $\phi_t(v)$ is the vector in SV which comes from parallel transporting the initial unit tangent vector along the unique unit speed geodesic passing through v at time zero). We shall use h to denote the topological entropy of ϕ.

In order to explain the coding we need to introduce two-sided subshifts of finite type. Given a zero-one matrix B, we can define the two-sided subshift of finite type $\sigma : X_B \to X_B$ on the space

$$X_B = \{x = (x_n)_{n=-\infty}^\infty \prod_{n=-\infty}^\infty \{1, \ldots, k\} \,:\, B(x_n, x_{n+1}) = 1, \forall n \in \mathbb{Z}\}$$

by $\sigma(x_n)_{n=-\infty}^\infty = (x_{n+1})_{n=-\infty}^\infty$. We can define a metric on X_B by

$$d((x_n)_{n=-\infty}^\infty, (y_n)_{n=-\infty}^\infty) = \sum_{n=-\infty}^\infty \frac{1 - \delta(x_n, y_n)}{2^{|n|}}.$$

With this metric X_B is compact and zero dimensional and $\sigma : X_B \to X_B$ is a homeomorphism.

For each $1 \le i \le k$ we denote $[i] = \{(x_n)_{n=-\infty}^{\infty} : x_0 = i\}$.

Given a strictly positive Hölder continuous function $\hat{r} : X_B \to \mathbb{R}^+$ we define the $\hat{r}$-suspension space by

$$X_B^{\hat{r}} = \{(x, u) : x \in X_B, 0 \le u \le \hat{r}(x)\}/(x, \hat{r}(x)) \sim (\sigma x, 0)$$

and a suspended flow $\sigma_t^{\hat{r}} : X_B^{\hat{r}} \to X_B^{\hat{r}}$ defined by $\sigma_t^{\hat{r}}(x, u) = (x, u + t)$ (subject to the identifications).

The next result relates the geodesic flow to a suspended flow.

Proposition 1. *There exists an aperiodic matrix B, a strictly positive Hölder continuous map $\hat{r} : X_B \to \mathbb{R}$ and a surjective continuous map $\pi : X_B \to SV$ such that*

(1) $\pi \circ \sigma_t^{\hat{r}} = \phi_t \circ \pi$
(2) *every closed ϕ-orbit corresponds to a σ^r-orbit of the same (prime) period, with at most finitely many exceptions*

(cf. [1], [15].*)*

Let $\mathcal{T} = \{T_1, \dots, T_k\}$ be the family of two dimensional local cross sections to the geodesic flow given by $\pi([i] \times \{0\})$, $i = 1, \dots, k$. The shift $\sigma : X_B \to X_B$ models the Poincaré return map $\mathcal{P} : \coprod_{i=1}^k T_i \to \coprod_{i=1}^k T_i$ and the function $\hat{r} : X_B \to \mathbb{R}$ models the return time $r : \coprod_{i=1}^k T_i \to \mathbb{R}$, in particular, $\phi_{r(x)}(x) = \mathcal{P}(x)$ (on the interiors of the disk).

Given $\epsilon > 0$ we can define the associated *local stable manifold* and *local unstable manifold* for a point $x \in SV$ by

$$W_\epsilon^{ss}(x) = \{y \in SV : d(\phi_t(x), \phi_t(y)) \le \epsilon, \forall t \ge 0\}$$

and

$$W_\epsilon^{su}(x) = \{y \in SV : d(\phi_{-t}(x), \phi_{-t}(y)) \le \epsilon, \forall t \ge 0\}.$$

Provided that $\epsilon > 0$ is sufficiently small these sets are diffeomorphic to one-dimensional embedded disks.

The local cross sections can be adjusted in the flow direction so that each $i = 1, \dots, k$:

(i) T_i is foliated by pieces of local stable manifolds $S_i(x) = W_\epsilon^{ss}(x) \cap T_i$; and
(ii) T_i contains a piece of local unstable manifold, which we denote by U_i.

This does not affect the underlying subshift of finite type $\sigma : X_B \to X_B$.

Remark. We can also define the *stable manifold* and *unstable manifold* for a point $x \in SV$ by

$$W^{ss}(x) = \{y \in SV : d(\phi_t(x), \phi_t(y)) \to 0, \text{ as } t \to +\infty\}$$

and

$$W^{su}(x) = \{y \in SV : d(\phi_{-t}(x), \phi_{-t}(y)) \to 0, \text{ as } t \to +\infty\}.$$

The families $\{W^{ss}(x) : x \in SV\}$ and $\{W^{ss}(x) : x \in SV\}$ are foliations of SV with one dimensional leaves. It is well-known that these foliations are not jointly integrable. In consequence, under assumption (ii), if $x \in T_i$ then its local unstable manifold $W^{su}_\epsilon(x)$ need not be contained in T_i.

Let $U_i(x)$ denote the projection of the local unstable manifold $W^{su}_\epsilon(x)$ along the flow lines onto T_i. The families $\{U_i(x) : x \in T_i\}$ and $\{S_i(x) : x \in T_i\}$, $i = 1, \ldots, k$, can be thought of as local stable and unstable manifolds for the Poincaré map.

For each $i = 1, \ldots, k$, the foliation $\{S_i(x) : x \in T_i\}$ gives a natural equivalence relation on the section T_i and we can identify the corresponding equivalence class I_i with U_i.

The Poincaré map $\mathcal{P} : \coprod_{i=1}^k T_i \to \coprod_{i=1}^k T_i$ gives rise to an associated interval map as follows. We project along the unstable manifolds $\{U_i(x)\}$ for each local cross section to obtain an induced expanding map $f : \coprod_{i=1}^k I_i \to \coprod_{i=1}^k I_i$. Moreover, we can choose $0 < \gamma < 1$ with $|f'(x)| \geq 1/\gamma$, for all $x \in \coprod_{i=1}^k I_i$ [16].

A desirable feature of this construction is that the function $r : \coprod_{i=1}^k T_i \to \mathbb{R}$ is constant on each unstable manifold $U_i(x)$. In consequence, there is an associated function $r : \coprod_{i=1}^k T_i \to \mathbb{R}$ for which we use the same notation. In particular, the function $\hat{r} = r \circ \pi : X_B \to \mathbb{R}$ depends only on the terms x_n, $n \geq 0$. Therefore, we can view $\hat{r}$ as a function defined on X_B^+.

1.3 Relating the two codings. The above two codings can be related by the following result.

Proposition 2. *We can choose the two sided subshift of finite type $\sigma : X_B \to X_B$ in Proposition 1 so that B is the matrix arising from the Bowen-Series coding.*

Proof. In constant curvature this is a reformulation of a theorem of Series [20]. For variable curvature this is an easy consequence of structural stability of Anosov flows and the connectedness of the space of all negatively curved metrics [6].

Let us recall how the directed graph arises in [Series]. The graph $\mathcal{G}$ is constructed geometrically by using the geometry of a fundamental domain to construct a partition of the boundary of the Poincaré disk into a finite number of intervals $I_1, \ldots, I_k$ and defining a map $f : I \to I$, where $I = \coprod_{i=1}^k I_i$. The vertices of $\mathcal{G}$ are identified with the intervals $I_1, \ldots, I_k$ and there is a directed edge from vertex i to vertex j if $f(I_i) \supset I_j$.

A sequence $x = (x_n) \in X_B$ is used to code a geodesic in the above proposition as follows. Each geodesic is uniquely determined by its two endpoints (realized as a pair of points on the boundary of the universal cover). The first point x_+ is coded using f in the sense that $f^n(x_+) \in I_{x_n}$ for $n \geq 0$. The second point is coded by a complementary interval map $g : J \to J$, where again $J = \coprod_{j=1}^k J_j$ is again a partition of the boundary of the Poincaré disk into a finite number of intervals $J_1, \ldots, J_k$ (in general different to $I_1, \ldots, I_k$). The reason that we get a two sided subshift of finite type $\sigma : X_B \to X_B$ is that the expanding map $f : I \to I$ has a natural extension to $\hat{f} : I \times J \to I \times J$ (i.e the projection onto the first co-ordinate is simply $f : I \to I$ and similarly, the projection onto the second co-ordinate of $\hat{f}^{-1} : I \times J \to I \times J$ is simply $g : J \to J$).

We want to realize the coding in [20,21] in terms of local cross-sections (as discussed in subsection 1.2). Given intervals I_i and J_j we can consider all geodesics on the universal cover $\mathbb{D}^2$ with end points in I_i and J_j. The tangent vectors to these geodesics form a subspace $\Omega \subset T_1\mathbb{D}^2$. If we consider a sufficiently large ball

$B(0,R) \subset \mathbb{D}^2$ then the intersection $\Omega \cap T_1 B(0,R)$ quotiented along the geodesic flow direction gives a two dimensional section $\hat{T}_{i,j}$. The only ambiguity in realizing this as a section in $T_1\mathbb{D}^2$ is how to place it in the flow direction. However, this is not a problem since $\hat{T}_{i,j}$ is only required as a section to the flow and thus its position can be varied under the geodesic flow.

The sections $\hat{T}_{i,j}$ constructed above are in $S\tilde{V}$. Each section $\hat{T}_{i,j}$ projects to a section $T_{i.j}$ for the geodesic flow on the compact manifold SV. We can carry out this construction for each of the pairs (i,j). The resulting family forms a Markov section for the geodesic flow by virtue of the properties $\hat{f}$.

2. Transfer Operators and zeta functions

Let us fix $s = \sigma + it$. We let $C^1(I)$ denote the Banach space of C^1 functions on I, with norm

$$||h||_{1,t} = \begin{cases} \max\left(||h||_\infty, \frac{||h'||_\infty}{|t|}\right) & \text{if } |t| \geq 1 \\ \max\left(||h||_\infty, ||h'||_\infty\right) & \text{if } |t| < 1 \end{cases}$$

Notice that with this (non-standard) weighting the norm depends on $|t|$.

Definition. Given $s \in \mathbb{C}$ we define the transfer operator $L_{-sr} : C^1(I) \to C^1(I)$ by

$$L_{-sr}w(x) = \sum_{fy=x} e^{-sr(y)} w(y)$$

Given a continuous function $g : I \to \mathbb{R}$ we define the *pressure* $P(g)$ by

$$P(g) = \sup\{h(\mu) + \int g d\mu \,:\, \mu = T\text{-invariant probability}\}.$$

In the special case that $t = 0$ the spectrum of $L_{-\sigma r} : C^1(I) \to C^1(I)$ is described by the following result.

Proposition 3 (Ruelle). *The operator $L_{-\sigma r}$ has maximal eigenvalue $e^{P(-\sigma r)}$ and the rest of the spectrum is contained in a ball of strictly smaller radius.*

The following result is due to Dolgopiat [3]. In particular, it gives estimates on the spectral radius of the operator $L_{-(\sigma+it)r}$ which are uniform in $t \in \mathbb{R}$. This will be of crucial importance in our analysis.

Proposition 4 (Dolgopiat, [3]). *There exist constants $\sigma_0 < h$, $C > 0$ and $0 < \rho < 1$ such that whenever $s = \sigma + it$ with $\sigma \geq \sigma_0$ and $n = p[\log|t|] + l$ ($p \geq 0$, $0 \leq l \leq [\log|t|] - 1$) then*

$$||L_{-sr}^n||_{1,t} \leq C\rho^{p[\log|t|]} e^{lP(-\sigma r)}$$

Corollary 4.1. *There exist constants $\sigma_0 < h$, $C > 0$ and $0 < \rho < 1$ such that whenever $s = \sigma + it$ with $\sigma \geq \sigma_0$ then the spectral radius of L_{-sr} is less than or equal to ρ.*

Remark. This result can also be extended to a corresponding statement for L_{-sr} acting on Hölder functions.

In order to study the counting function $\pi(T)$ we will consider a function of a complex variable called the zeta function $\zeta(s)$. The required estimates on $\pi(T)$ are obtained by analyzing the analytic domain of $\zeta(s)$.

The function $\zeta(s)$ is defined by the infinite product

$$\zeta(s) = \prod_{\gamma} \left(1 - e^{-sl(\gamma)}\right)^{-1},$$

where $s \in \mathbb{C}$ and γ denotes a closed (directed) geodesic on V of length $l(\gamma)$. This converges to a non-zero analytic function for $Re(s) > h$ (cf. [PP]).

The main properties of $\zeta(s)$ that we shall use are contained in the following proposition. The rest of this section will be devoted to its proof.

Proposition 5. *There exists $c_0 < h$ such that $\zeta(s)$ is analytic in the half-plane $Re(s) > c_0$, except for a simple pole at $s = h$. Moreover, there exists $0 < \alpha < 1$ such that*

$$\zeta'(\sigma + it)/\zeta(\sigma + it) = O(|t|^{\alpha}), \qquad as\ |t| \to +\infty$$

uniformly for $\sigma > c_0$.

In order to study $\zeta(s)$ it is convenient to re-express it in terms of the expanding map $f : I \to I$. We shall use the notation $Z_n(-sr) = \sum_{f^n x = x} e^{-sr^n(x)}$. We then have the following result (cf.[1], [11]).

Lemma 1.

$$\zeta(s) = \exp \sum_{n=1}^{\infty} \frac{1}{n} Z_n(-sr).$$

We use this formulation of the zeta function $\zeta(s)$ to relate it to the Ruelle transfer operator $L_{-sr} : C^1(I) \to C^1(I)$. In particular, we have the following result.

Lemma 2. *For each interval I_i we fix a point $x_i \in I_i$. For any $max\{\rho, \gamma\} < \rho_0 < 1$ there exists $C_1 > 0$ such that*

$$\left| Z_n(-sr) - \sum_{i=1}^{k} L_{-sr}^n \chi_{I_i}(x_i) \right| \leq C_1 |t| n \rho_0^n \tag{2.1}$$

where χ_{I_i} is the characteristic function for I_i.

Proof. This result is essentially proved in [18]. We recall the main ideas for the reader's benefit. To explain the ideas, assume that for each $n \geq 0$ we consider all strings $\underline{i} = (i_0, \ldots, i_{n-1})$ with $I_{\underline{i}} = I_{i_0} \cap f^{-1} I_{i_1} \cap \ldots \cap f^{-(n-1)} I_{i_{n-1}} \neq \emptyset$ and write $|\underline{i}| = n$.

We can consider the characteristic functions $\chi_{I_{\underline{i}}}$. This is not in $C^1(I)$, but it is an element of $C^1\left(\coprod_{|\underline{i}|=n} I_{\underline{i}}\right) = \oplus_{|\underline{i}|=n} C^1\left(I_{\underline{i}}\right)$. Observe that for $1 \leq m \leq n-1$ we have $L_{-\sigma r} : \oplus_{|\underline{i}|=m+1} C^1\left(I_{\underline{i}}\right) \to \oplus_{|\underline{i}|=m} C^1\left(I_{\underline{i}}\right)$. Thus, in particular, if $|\underline{i}| = m$ then $L^k_{-\sigma r}\chi_{I_{\underline{i}}} \in C^1(I)$.

We can choose $x_{\underline{i}} \in I_{\underline{i}}$ such that:

(1) $f^n x_{\underline{i}} = x_{\underline{i}}$ if possible; and
(2) $x_{\underline{i}}$ is arbitrary otherwise.

If $\underline{i} = (i_0, \ldots, i_{m-1})$ then we can adopt the notation $\underline{j}(\underline{i}) = (i_0, \ldots, i_{m-2})$ (i.e. we look at the shorter string after throwing away the last term).

From these observations, we see that

$$Z_n(-sr) = \sum_{|\underline{i}|=n} (L^n_{-sr}\chi_{I_{\underline{i}}})(x_{\underline{i}}).$$

In addition, we can write

$$\begin{aligned}
&\sum_{|\underline{i}|=n} L^n_{-sr}\chi_{I_{\underline{i}}}(x_{\underline{i}}) - \sum_{i=1}^{k} L^n_{-sr}\chi_{I_i}(x_i) \\
&= \sum_{m=1}^{n}\left(\sum_{|\underline{i}|=m} L^n_{-sr}\chi_{I_{\underline{i}}}(x_{\underline{i}}) - \sum_{|\underline{j}|=m-1} L^n_{-sr}\chi_{I_{\underline{j}}}(x_{\underline{j}})\right) \\
&= \sum_{m=1}^{n}\left(\sum_{|\underline{i}|=m}\left(L^n_{-sr}\chi_{I_{\underline{i}}}(x_{\underline{i}}) - L^n_{-sr}\chi_{I_{\underline{i}}}(x_{\underline{j}(\underline{i})})\right)\right).
\end{aligned}$$

Observe that many of the terms in these summations may well be zero and others may be very simple. For example

$$L^n_{-sr}\chi_{I_{\underline{i}}}(x_{\underline{j}(\underline{i})}) = e^{-sr^n\left((f|I_{\underline{i}})^{-1}x_{\underline{j}(\underline{i})}\right)}.$$

Since for $|\underline{i}| = m$ we have that $L^m_{-sr}\chi_{I_{\underline{i}}} \in C^1(I)$ we can use the estimate

$$\begin{aligned}
&\left|\sum_{|\underline{i}|=n} L^n_{-sr}\chi_{I_{\underline{i}}}(x_{\underline{i}}) - \sum_{i=1}^{k} L^n_{-sr}\chi_{I_i}(x_i)\right| \\
&\leq \sum_{m=1}^{n-1} ||L^{n-m}_{-sr}||_{C^1} \sum_{|\underline{i}|=m} ||L^m_{-sr}\chi_{I_{\underline{i}}}||_{C^1} d(x_{\underline{i}}, x_{\underline{j}(\underline{i})}).
\end{aligned} \tag{2.2}$$

We can estimate

(1) $||L^{n-m}_{-\sigma r}||_{C^1} \leq C\rho^{k[\log|t|]}e^{lP(-\sigma r)} \leq C\rho_0^{k[\log|t|]}e^{lP(-\sigma r)}$, by Proposition 4
(2) $||L^m_{-sr}\chi_{I_{\underline{i}}}||_{C^1} \leq \text{const.}|t|$, since $L^m_{-sr}\chi_{I_{\underline{i}}} = e^{-sr^m\left((f|I_{\underline{i}})^{-1}x_{\underline{j}(\underline{i})}\right)}$,
(3) $d(x_{\underline{i}}, x_{\underline{j}(i)}) \leq \text{const.}\gamma^m \leq \text{const.}\rho_0^m$ where $|\underline{i}| = m$ and $|f'| \geq 1/\gamma > 1$

We can compare these estimates with the inequality (2.2) to deduce inequality (2.1). This completes the proof of Lemma 2.

We now return to the proof of Proposition 5. Let us fix $\rho_0 > \rho$ (in Lemma 2). Using Lemma 2 and Proposition 4 we can bound

$$\begin{aligned} |Z_n(-sr)| &\leq \left| Z_n(-sr) - \sum_{i=1}^{k} L_{-sr}^n \chi_{I_i}(x_i) \right| + \left| \sum_{i=1}^{k} L_{-sr}^n \chi_{I_i}(x_i) \right| \\ &\leq C_1 |t| n \rho_0^n + \sum_{i=1}^{k} ||L_{-sr}^n|| \\ &\leq C_1 |t| n \rho_0^n + k \left(C \rho_0^{p[\log t]} e^{lP(-\sigma r)} \right), \end{aligned}$$

where $n = p[\log |t|] + l$.

This gives us the following estimate on $\log |\zeta(s)|$ for $s = \sigma + it$ where $\sigma > \sigma_0$ and $|t| \geq 1$:

$$\begin{aligned} \log |\zeta(s)| &\leq \sum_{n=1}^{\infty} \frac{1}{n} |Z_n(-sr)| \\ &\leq \sum_{n=1}^{\infty} C_1 |t| \rho_0^n + kC \sum_{p=0}^{\infty} \rho^{p[\log |t|]} \left(\sum_{l=0}^{[\log |t|]-1} e^{lP(-\sigma r)} \right) \\ &\leq C_1 |t| + kC \sum_{p=0}^{\infty} \rho^{p[\log |t|]} \left(\sum_{l=0}^{[\log |t|]-1} e^{lP(-\sigma r)} \right) \\ &\leq C_1 |t| + kC \left(\frac{1}{1 - |t|^{-|\log \rho|}} \right) \max \left\{ [\log |t|], \frac{e^{[\log |t|] P(-\sigma r)} - 1}{e^{P(-\sigma r)} - 1} \right\}. \end{aligned}$$

For $|t|$ sufficiently large we see that $1/\left(1 - |t|^{-|\log \rho|}\right)$ is uniformly bounded and thus

$$\log |\zeta(s)| \leq C_1 |t| + C_2 \max \left([\log |t|], |t|^{P(-\sigma r)} \right), \tag{2.3}$$

for some $C_2 > 0$.

To proceed, we need a standard result from complex analysis which allows us to convert (2.3) into a bound for $\zeta'(s)/\zeta(s)$ in the half-plane $\sigma > \sigma_0$.

Lemma 3 ([4, Theorem 4.2]). *Let $z \in \mathbb{C}$. Given $R > 0$ and $\epsilon > 0$ suppose that $F(s)$ is analytic on the disk $\Delta = \{s = \sigma + it : |s - z| \leq R(1 + \epsilon)^3\}$ and that there are no zeros for $F(s)$ on the open subset*

$$\{s = \sigma + it \in \mathbb{C} : |s - z| \leq R(1 + \epsilon)^2 \text{ and } \sigma > Re(z) - R(1 + \epsilon)\}.$$

Suppose in addition that there exists a constant $U(z) \geq 0$ such that $\log |F(s)| \leq U(z) + \log |F(z)|$ on the set $\{s = \sigma + it : |s - z| \leq R(1 + \epsilon^3)\}$. Then we have the following bound for the logarithmic derivative on the disk $\{s = \sigma + it : |s - z| \leq R\}$:

$$\left| \frac{F'(s)}{F(s)} \right| \leq \frac{2 + \epsilon}{\epsilon} \left(\left| \frac{F'(z)}{F(z)} \right| + \frac{\left(2 + \frac{1}{(1+\epsilon)^2}\right)(1 + \epsilon)}{R\epsilon^2} U(z) \right).$$

Figure 1. Applying the Lemma

We want to apply the above lemma to $\zeta(s)$ where $s = \sigma + it$ lies in the strip $\sigma_1 < \sigma \leq h$, where $\sigma_1 > \sigma_0$. (We need to consider this smaller strip in order to get a uniform bound on the logarithmic derivative.)

We set $z = (h+1) + it$. In particular, this implies that we have the uniform lower bound $|\zeta(z)| \geq \frac{1}{|\zeta(Re(z))|} \geq \frac{1}{|\zeta(h+1)|}$, independent of t. In consequence, we see that by (2.3) we can choose $U(z)$ in Lemma 3 to satisfy

$$U(z) \leq C_3|t| + C_4 \max\{\log|t| + |t|^{P(-\sigma_1 r)}\}.$$

If we set $R = 1 + \frac{(h-\sigma_1)}{2}$ then we can choose $\epsilon > 0$ such that $R(1+\epsilon)^3 = 1 + (h - \sigma_1)$. In particular, Δ is contained in the half-plane $Re(s) > \sigma_1$ and thus is a non-zero analytic domain for $\zeta(s)$.

This allows us to deduce that there exists a constant $C_5 > 0$ such that for $Re(s) > \sigma_2 := \frac{h+\sigma_1}{2}$ we have

$$\left|\frac{\zeta'(s)}{\zeta(s)}\right| \leq C_5 \max\{|t|, |t|^{P(-\sigma_1 r)}\}.$$

Furthermore, since $P(-hr) = 0$ and $\sigma \mapsto P(-\sigma r)$ is continuous [17] we can assume without loss of generality that $P(-\sigma_1 r) \leq 1$ by choosing σ_1 sufficiently close to h.

To complete the proof we need to show that we can choose the exponent of $|t|$ to be strictly less than 1. To do this we make use of the well-known Phragmén-Lindelöf Theorem [22, §5.65]. We know that the function $(\zeta'/\zeta)(s) + (s-h)^{-1}$ is analytic in the strip $\sigma_2 \le Re(s) \le h+\delta$, for any $\delta > 0$. Moreover, we have the bounds

(1) $|(\zeta'/\zeta)(\sigma_2 + it)| \le C_5|t|$
(2) for each $\delta > 0$ there exists $C_6 > 0$ such that $|(\zeta'/\zeta)(h+\delta+it)| \le C_6$, (cf. [11]).

The Phragmén-Lindelöf Theorem tell us that for any $\sigma_2 \le \sigma_3 \le h+\delta$ we have that $|(\zeta'/\zeta)(\sigma_3 + it)| = O(|t|^\alpha)$, where

$$\alpha = \left(\frac{(h+\delta) - \sigma_3}{(h+\delta) - \sigma_2} \right).$$

Provided we fix $\sigma_3 > \sigma_2$ then $0 < \alpha < 1$.

By taking $c_0 = \sigma_3$, the proof of Proposition 5 is complete.

4. The error term in counting closed orbits

In this section we shall establish the estimates that prove Theorem 1. To prove the asymptotic estimate it is technically easier to establish first the corresponding result for the following related functions.

For $T > 0$ define $\psi(T) = \sum_{e^{nhl(\gamma)} \le T} hl(\gamma)$ and $\psi_1(T) = \int_1^T \psi(x)dx$. It is an easy observation that $\psi_1(T) = \sum_{e^{nhl(\gamma)} \le T} hl(\gamma)(T - e^{nhl(\gamma)})$.

It will prove convenient to slightly modify the original definition of the zeta function and instead work with the function

$$\zeta_0(s) = \prod_\gamma \left(1 - e^{-shl(\gamma)}\right)^{-1}.$$

Observe that $\zeta(sh) = \zeta_0(s)$. Clearly similar conclusions to those of Proposition 5 for $\zeta(s)$ also hold for $\zeta_0(s)$. However, for simplicity, let us assume that $\zeta_0(s)$ is analytic (with the corresponding bounds) in the strip $Re(s) > c_0$, for some $c_0 < 1$.

Using our estimates on the logarithmic derivative of the zeta function from the previous section we obtain the following elegant formula for $\psi_1(t)$.

Proposition 6. *If $c_0 < c < 1$ then*

$$\psi_1(T) = \frac{T^2}{2} + \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left(-\frac{\zeta_0'(s)}{\zeta_0(s)} \right) \frac{T^{s+1}}{s(s+1)} ds.$$

In particular, $\psi_1(T) = \frac{T^2}{2} + O(T^{c+1})$.

Proof. One need only apply the identity

$$\frac{1}{2\pi i} \int_{d-i\infty}^{d+i\infty} \frac{y^{s+1}}{s(s+1)} ds = \begin{cases} 0 \text{ if } 0 < y \le 1 \\ y-1 \text{ if } y > 1 \end{cases}$$

[4, p. 50] with $d > 1$ term by term to $-\zeta_0'(s)/\zeta_0(s) = \sum_{n=1}^{\infty} \sum_\gamma hl(\gamma) e^{-snhl(\gamma)}$ to obtain

$$\psi_1(T) = \frac{1}{2\pi i} \int_{d-i\infty}^{d+i\infty} \left(-\frac{\zeta_0'(s)}{\zeta_0(s)} \right) \frac{T^{s+1}}{s(s+1)} ds$$

Using the estimate $\zeta'(s)/\zeta(s) = O(|t|^\alpha)$ we can see that $\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left(-\frac{\zeta_0'(s)}{\zeta_0(s)} \right) \frac{T^{s+1}}{s(s+1)} ds$ exists and satisfies the bound

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left(-\frac{\zeta_0'(s)}{\zeta_0(s)} \right) \frac{T^{s+1}}{s(s+1)} ds = O\left(T^{c+1}\right)$$

Furthermore,

$$\begin{aligned} &\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left(-\frac{\zeta_0'(s)}{\zeta_0(s)} \right) \frac{T^{s+1}}{s(s+1)} ds - \frac{1}{2\pi i} \int_{d-i\infty}^{d+i\infty} \left(-\frac{\zeta_0'(s)}{\zeta_0(s)} \right) \frac{T^{s+1}}{s(s+1)} ds \\ &= \lim_{R \to +\infty} \frac{1}{2\pi i} \int_{\Gamma_R} \left(-\frac{\zeta_0'(s)}{\zeta_0(s)} \right) \frac{T^{s+1}}{s(s+1)} ds \end{aligned}$$

where Γ_R is the contour consisting of line segments joining $d - iR$, $d + iR$, $c + iR$ and $c - iR$. The result now follows by applying the residue theorem to see that

$$\frac{1}{2\pi i} \int_{\Gamma_R} \left(-\frac{\zeta_0'(s)}{\zeta_0(s)} \right) \frac{T^{s+1}}{s(s+1)} ds = \frac{T^2}{2}$$

for all $R > 0$.

We can now use elementary arguments to translate the estimate for $\psi_1(T)$ into an estimate for $\psi(T)$.

Proposition 7. $\psi(T) = T + O(T^{(c+1)/2})$.

Proof. We define a function of T by $\epsilon = \epsilon(T) := T^{(c+1)/2}$. Since $T \to \psi(T)$ is monotone increasing we see that

$$\frac{\psi_1(T+\epsilon) - \psi_1(T)}{\epsilon} = \frac{1}{\epsilon} \int_T^{T+\epsilon} \psi(t) dt \geq \psi(T).$$

Observe that by Proposition 6

$$\begin{aligned} \frac{\psi_1(T+\epsilon) - \psi_1(T)}{\epsilon} &= \frac{1}{\epsilon} \left(\frac{(T+\epsilon)^2}{2} - \frac{T^2}{2} + O\left(T^{c+1}\right) \right) \\ &= T + O\left(\epsilon, \frac{T^{c+1}}{\epsilon} \right) \\ &= T + O\left(T^{(c+1)/2} \right) \end{aligned}$$

Thus $\psi(T) \leq T + O\left(T^{(c+1)/2}\right)$. A similar argument based on the inequality

$$\frac{\psi_1(T) - \psi_1(T-\epsilon)}{\epsilon} = \frac{1}{\epsilon} \int_{T-\epsilon}^{T} \psi(t) dt \leq \psi(T).$$

gives that $\psi(T) \geq T + O\left(T^{(c+1)/2}\right)$. This complete the proof.

We are now in a position to complete the proof of Theorem 1. We define $\pi_0(T) = \sum_{e^{nhl(\gamma)} \leq T} 1$ and observe that

$$\begin{aligned}\pi_0(T) &= \int_2^T \frac{1}{\log x} d\psi(x) + O(1) \\ &= \left[\frac{\psi(x)}{\log x}\right]_2^T + \int_2^T \psi(x) \frac{d}{dx}\left(-\frac{1}{\log x}\right) dx + O(1).\end{aligned}$$

However the identity

$$\int_2^T x \frac{d}{dx}\left(-\frac{1}{\log x}\right) dx + \frac{T}{\log T} = \mathrm{li}(T) + \frac{2}{\log 2}$$

allows us to conclude that $\pi_0(T) = \mathrm{li}(T) + O\left(T^{(c+1)/2}/\log T\right)$. We introduce $\pi_1(T) = \sum_{e^{hl(\gamma)} \leq T} 1$. Clearly,

$$\pi_0(T) = \pi_1(T) + \sum_{n \geq 2} \pi(T^{1/n}) = \pi_1(T) + O\left(T^{1/2} \log T\right).$$

In particular, $\pi_1(T) = \mathrm{li}(T) + O\left(T^{(c+1)/2}/\log T\right) + O\left(T^{1/2}\log T\right) = O(T^{c'})$, for any c' satisfying $\max\{1/2, (c+1)/2\} < c' < 1$. By changing variables we obtain

$$\pi(T) = \mathrm{li}(e^{hT}) + O\left(e^{c'hT}\right),$$

where we observe that $c'h < h$. This proves the theorem.

5. Poincaré series

We now turn our attention to the proof of Theorem 2. In order to obtain the promised estimate on $N(T)$ we need to study an appropriate analytic function. This function is the Poincaré series

$$\eta(s) = \sum_{g \in \Gamma} e^{-sd(x,gx)}$$

(which converges to an analytic function for $Re(s) > h$).

The following proposition allows us to write the Poincaré series in a more convenient symbolic form.

Proposition 8. *There exists a Hölder continuous function $R : X_A^+ \to \mathbb{R}$ such that:*

(a) *for every allowed word $(*, i_1, \ldots, i_n)$ with associated group element $g = \lambda(*, i_1) \ldots \lambda(i_{n-1}, i_n)$ we have that $d(x, gx) = R^n(*, i_1, \ldots, i_n, 0, 0, \ldots)$ where $R^n(x) = R(x) + R(\sigma x) + \ldots + R(\sigma^{n-1}x)$;*

(b) *for every closed loop $(i_0, i_1, \ldots, i_n)$ in $\mathcal{G}$ with $i_0 = i_n$ and with associated group element $g = \lambda(i_0, i_1) \ldots \lambda(i_{n-1}, i_n)$ we have that*

$$l(\gamma) = R^n(i_0, i_1, \ldots, i_{n-1}, i_0, i_1, \ldots, i_{n-1}, \ldots),$$

where γ is the unique closed geodesic in the conjugacy class of g and where $l(\gamma)$ denotes its length;

(c) *the restriction $R : X_B^+ \to \mathbb{R}^+$ and the function $\hat{r} : X_B^+ \to \mathbb{R}^+$ differ by a Hölder coboundary (i.e, there exists a continuous function $u : X_B^+ \to \mathbb{R}$ such that $R = \hat{r} + u \circ \sigma - u$).*

Part (a) was established by the authors in [14]. Part (b) has essentially established in [14, Proposition 2]. Part (c) follows from part (b) and a consideration of the closed orbits of the geodesic flow and an application of Livsic's theorem [9].

Using Proposition 8 we can rewrite $\eta(s)$ in terms of the subshift of finite type as

$$\eta(s) = 1 + \sum_{n=1}^{\infty} \sum_{z \in S_n} e^{-sR^n(z\dot{0})},$$

where S_n denotes the set of all allowed paths $z = z_0 \dots z_n$ in $\mathcal{G}$ of (edge) length n with $z_0 = *$ and where $z\dot{0} = (z_0, \dots, z_n, 0, 0, \dots)$ (cf. [12], [13], [14]).

The complex function $\eta(s)$ can be studied via a modified transfer operator.

Definition. We define $M_{-sR} : C^\alpha(X_A^+) \to C^\alpha(X_A^+)$ by

$$M_{-sR}g(x) = \sum_{\substack{\sigma y = x \\ y \neq \dot{0}}} e^{-sR(y)} g(y),$$

where we equip $C^\alpha(X_A^+)$ with the norm

$$||g||_{\alpha,t} = \begin{cases} \max\left(||g||_\infty, \frac{||g||_\alpha}{|t|}\right) & \text{if } |t| \geq 1 \\ \max\left(||g||_\infty, ||g||_\alpha\right) & \text{if } |t| < 1 \end{cases}.$$

Here we let $|h|_\alpha = \sup_{x \neq y} \frac{|h(x)-h(y)|}{d(x,y)^\alpha}$ and t be the imaginary part of s.

Using this definition we may rewrite the Poincaré series in the form

$$\eta(s) = 1 + \sum_{n=1}^{\infty} M_{-sR}^n(\chi_{[*]})(\dot{0})$$

where $\chi_{[*]}$ is the indicator function for the set of sequences in X_A^+ beginning with $*$.

We need to present an analogue to Proposition 4 giving a bounds on $||M_{-sR}^n||$. This takes the following form.

Proposition 9. *There exists constants $C > 0$, $\sigma_0 < h$ and $0 < \rho < 1$ such that whenever $s = \sigma + it$ with $\sigma_0 \geq \sigma$ and $n = p[\log|t|] + l$ ($p \geq 0$, $0 \leq l \leq [\log|t|] - 1$) then*

$$||M_{-sR}^n|| \leq C\rho^{k[\log|t|]} e^{lP(-\sigma R)}$$

We postpone the proof of this result until the next section.

We now use Proposition 9 to estimate $\eta(s)$. For $Re(s) > \sigma_0$ and $|t| \geq 1$ we have

$$\begin{aligned} |\eta(s)| &\leq \sum_{k=0}^{\infty} \sum_{l=1}^{[\log|t|]-1} C\rho^{k[\log|t|]} e^{lP(-\sigma R)} \\ &\leq \frac{1}{1-\rho^{[\log|t|]}} \sum_{l=0}^{[\log|t|]-1} Ce^{lP(-\sigma R)} \\ &\leq C' \max\left\{\log|t|, |t|^{P(-\sigma R)}\right\}. \end{aligned}$$

Observe, that without loss of generality, we may choose σ_0 sufficiently close to h to ensure that $P(-\sigma_0 R) < 1$ (this again follows by continuity of $\sigma \mapsto P(-\sigma R)$ and the fact that $P(-hR) = 0$).

This estimate allows us to conclude the following.

Proposition 10. *There exists $c_0 < h$ such that $\eta(s)$ is analytic in the half-plane $Re(s) > c_0$, except for a simple pole at $s = h$. Moreover, writing $s = \sigma + it$, there exists $0 < \alpha < 1$ and $C'' > 0$ such that for $c_0 < \sigma < h$ we have*

$$|\eta(s)| \leq C''|t|^\alpha, \quad \text{for } |t| \geq 1.$$

Remark. Notice that, in contrast to the case of Proposition 5, the proof of Proposition 10 does not require the Phragmén-Lindelöf Theorem.

6. Norm Estimates

In this section we give the proof of Proposition 9. The argument follows the same lines as that in [3], although during the proof we need to address three important differences between the operators M_{-sR} and L_{-sr}:

(a) M_{-sR} acts on Hölder functions whereas L_{-sr} acts on C^1 functions;
(b) $\sigma : X_A^+ \to X_A^+$ is not transitive, whereas $f : I \to I$ here are two cases to consider and $\sigma : X_B^+ \to X_B^+$ are transitive; and
(c) we replace r with R.

6.1 Generalities. In this subsection we explain the salient features of the geodesic flow which lead to the norm estimates. Recall that the expanding map $f : \coprod_{i=1}^k I_i \to \coprod_{i=1}^k I_i$ is modelled by the mixing one-sided subshift $\sigma : X_B^+ \to X_B^+$. We fix $n_0 > 0$ such that for the corresponding aperiodic transition matrix B we have that B^{n_0} has all entries positive. Given $n \geq n_0$ we write for $y_1, y_2 \in I$ that $y_1 \sim_n y_2$ if there exists $1 \leq i \leq k$ with $f^n(y_1), f^n(y_2) \in I_i$ and $f^n(y_1) = f^n(y_2)$. Given $x \in I$ we shall:

(1) consider distinct $y_1 \sim_n y_2$ such that $x = f^n(y_1) = f^n(y_2)$.

We can fix a $x^0 \in I$ and

(2) choose y_1^0, y_2^0 (corresponding to y_1, y_2, respectively) such that $x^0 = f^n(y_1^0) = f^n(y_2^0)$

We introduce the function

$$\psi(x, x^0) = (r^n(y_1) - r^n(y_2)) - \left(r^n(y_1^0) - r^n(y_2^0)\right)$$

An essential feature of the stable and unstable foliations for the geodesic flow is their uniform non-integrability [2], [5], which is associated to the contact property. In particular, it allows for each x^0 (and sufficiently close x) choices of y_1, y_2, y_1^0, y_2^0 as in (1) and (2) above such that the maps $I_i \ni x \mapsto \psi(x, x^0)$ are C^1 and locally strictly monotonic. Furthermore, there exists constants $B_1, B_2 > 0$ with

$$B_1|x - x^0| \leq |\psi(x, x^0)| \leq B_2|x - x^0|.$$

FIGURE 2. Non joint integrability.

6.2 Relating C^α and C^0 norms. Our main goal is to show that the operator $M_{-(\sigma+it)R}$ is a contraction with respect to the norm $||\cdot||_{\alpha,t}$ on $C^\alpha(X_A^+)$. In this subsection we introduce a proposition which will allow us to achieve this contraction via the a priori weaker $||\cdot||_\infty$ contraction of the same operator.

For the purposes of the proof the function r should, in fact, depend on the value of σ. We shall denote this by r_σ. More specifically, by adding coboundaries (depending on σ) we can arrange that, for $\sigma_0 \leq \sigma \leq h$, we have $L_{-\sigma r_\sigma}1 = e^{P(-sigmar_\sigma)}1$. In particular, $r_h = r$ and $L_{-hr_h}1 = 1$ [11]. By adding coboundaries (depending on σ) we can assume that for each $\sigma_0 \leq \sigma \leq h$ the roof function r_σ satisfies $L_{-\sigma r_\sigma}1 = e^{P(-\sigma r_\sigma)}1$.

By Proposition 8(c) we know that R_h restricted to X_B^+ is cohomologous to $\hat{r}_h$. Similarly, we see that R_σ restricted to X_B^+ is cohomologous to $\hat{r}_\sigma$. The following result is essentially a well-known estimate [11] (observing from the proof that no transitivity assumption on the shift is required).

Proposition 11. *There exists $C > 1$ and $0 < \theta < 1$ (both independent of $|t| \geq 1$ and $\sigma_0 \leq \sigma \leq h$) such that*

$$||(L^n_{-(\sigma+it)r_\sigma}w)'||_\infty \leq e^{nP(-\sigma r_\sigma)}\left(C|t|||w||_\infty + \theta^n||w'||_\infty\right), \qquad \textit{for } n \geq 0$$

and

$$|M^n_{-(\sigma+it)R_\sigma} g|_\alpha \leq ||M^n_{-\sigma R_\sigma}||_\infty \left(C|t|||g||_\infty + \theta^n |g|_\alpha\right), \quad \textit{for } n \geq 0.$$

For simplicity, we will ignore the contributions of $e^{P(-\sigma r_\sigma)}$ and $||M^n_\sigma||_\infty$ (since these can be made to grow at arbitrarily slow rates by choosing σ sufficiently close to 1). In order to actually show that $M_{-(\sigma+it)R_\sigma}$ is a $||\cdot||_{\alpha,t}$ contraction we have consider two cases: the first (relatively easy) case is when $2C|t|||g||_\infty \leq |g|_\alpha$; the second is when $2C|t||g|_\infty \geq |g|_\alpha$. We consider these two situations separately.

6.3 Case I : $2C|t|||g||_\infty \leq |g|_\alpha$**.**

If we make the additional hypothesis that $2C|t|||g||_\infty \leq |g|_\alpha$ then we can fix $\frac{1}{2} < \eta < 1$ and choose $k > 0$ and $\sigma_0 < h$ sufficiently large that $\left(\frac{1}{2} + \theta^k\right) < \eta$. Proposition 11 gives that

$$\frac{1}{|t|}|M^k_{-(\sigma+it)R_\sigma} g|_\alpha \leq C||g||_\infty + \theta^k \frac{1}{|t|}|g_\alpha \leq \left(\frac{1}{2} + \theta^k\right) \frac{1}{|t|}|g|_\alpha \leq \eta ||g||_{\alpha,t}.$$

In addition, $||M^k_{-(\sigma+it)R_\sigma} g||_\infty \leq ||g||_\infty \leq \frac{1}{2C}\frac{1}{|t|}|g|_\alpha \leq \eta||g||_{\alpha,t}$. Together these two inequalities show that $||M^k_{-(\sigma+it)R_\sigma} g||_{\alpha,t} \leq \eta||g||_{\alpha,t}$.

6.4 Case II : $|g|_\alpha \leq 2C|t|||g||_\infty$**.** To prove $||\cdot||_\infty$-convergence (and consequently $||\cdot||_{\alpha,t}$-convergence) we first establish $L^1(\mu_\sigma)$-convergence with respect to an appropriate measure μ_σ.

A key observation is that the measure μ_σ is supported on $X_B^+ \subset X_A^+$. The first important consequence is that we shall only need to consider the corresponding operator on functions defined on the transitive shift space X_B^+. The second important feature is that for L^1 contraction we need only study $L_{-(\sigma+it)r_\sigma}$ rather than $M_{-(\sigma+it)R_\sigma}$ because of the fact that R and $\hat{r}$ are cohomologous on X_B^+ by Proposition 8(c).

Let us assume for simplicity that g is induced from a function $\hat{w} \in C^1(I)$ (i.e. $g = w \circ \pi$ under the canonical mapping $\pi : X_B^+ \to I$) and that $w \in C^1(I)$ satisfies $||\hat{w}'||_\infty \leq 2C|t|||w||_\infty$ and $||\hat{w}||_\infty = 1$ then we want to associate a sequence of functions $u_N > 0$, $N \geq 0$, such that

(1) $0 \leq |L^{nN}_{-(\sigma+it)r_\sigma} g(x)| \leq u_N(x)$;

(2) There exists $0 < \beta < 1$ such that $\int u_N(x) d\mu(x) \leq \beta^n$ (and β is independent of $\hat{h}$, t and σ).

In addition, the functions u_N are constructed so that they are C^1 on each U_i and

(3) $|\log u_N'(x)| = |\frac{u_N'(x)}{u_N(x)}| \leq 2C|t|$.

Remark. For completeness, we mention that the functions u_N are defined inductively as follows:

(i) Fix $u_0 = 1$;

(ii) Given the C^1 function $u_N(x)$ we want to pair up corresponding terms y_1, y_1 for $L^n_{-(\sigma+it)r_\sigma} u_N$. We can choose $0 < \eta_0 < 1$, $0 \leq \theta_0 \leq 2\pi$ and $\frac{\pi}{2} > \delta > 0$ such that whenever $t|r^n(y_1) - r^n(y_2)| \in [\theta_0 - \delta, \theta_0 + \delta] \pmod{2\pi}$ then

$$\begin{aligned} &|e^{-(\sigma+it)r^n(y_1)} h(y_1) + e^{-(\sigma+it)r^n(y_1)} h(y_2)| \\ &\leq \eta \left(e^{-\sigma r^n(y_1)} u_n(y_1) + e^{-\sigma r^n(y_2)} u_n(y_2)\right). \end{aligned}$$

The variation of $x \mapsto u_n(y), h(y)$ (and their arguments) can be made arbitrarily small by arranging m sufficiently large. Let

$$\mathcal{A}_{t,\theta_0,\delta} = \{x : t\psi(w_1, w_2) \in [\theta_0 - \delta, \theta_0 + \delta] \pmod{2\pi}\}$$

then we can choose a smooth function $\eta_0 \le \eta(y) \le 1$ such that

$$\eta(y) = \begin{cases} \eta_0 & \text{if } x \in \mathcal{A}_{t,\theta_0,\frac{\delta}{2}} \\ 1 & \text{if } x \in \left(\cup_{i=1}^k U_i\right) - \mathcal{A}_{t,\theta_0,\delta} \end{cases}.$$

We then set $u_{N+1}(x) = L^{n+m}_{-\sigma r_\sigma}(\eta u_N)(x)$.

The following lemma gives important estimates on the probability measures satisfying $L^*_{-\sigma r_\sigma}\mu_\sigma = \mu_\sigma$.

Lemma 4. *If we identify U_i with $[0,1]$ then we write $\mathcal{A}_{t,\theta_0,\frac{\delta}{2}} = \cup_{i=1}^k [t_{2i}, t_{2i+1}]$ where $0 \le t_1 < t_2 < \ldots < t_{2k+1} \le 1$, say, then for all $\sigma_0 < \sigma < h$:*

(a) *There exists $C_1, C_2 > 0$ such that $C_1 \le \frac{\mu_\sigma([t_{2i+1}, t_{2i+2}])}{\mu_\sigma([t_{2i}, t_{2i+1}])} \le C_2$;*

(b) *There exists $C_3, C_4 > 0$ such that $C_3 \le \frac{u_n(z')}{u_n(z)} \le C_4$ for $t_{2i} \le z \le t_{2i+1}$ and $t_{2i+1} \le z' \le t_{2i+2}$ for all $n \ge 0$; and*

(c) *There exists $0 < \alpha < 1$ such that*

$$\int_{\mathcal{A}_{t,\theta_0,\frac{\delta}{2}}} u_{n+1}(x) d\mu_\sigma \le \alpha \int_{\mathcal{A}_{t,\theta_0,\frac{\delta}{2}}} u_n(x) d\mu_\sigma, \qquad \textit{and}$$

$$\int_{\mathcal{A}^c_{t,\theta_0,\frac{\delta}{2}}} u_{n+1}(x) d\mu_\sigma \le \int_{\mathcal{A}^c_{t,\theta_0,\frac{\delta}{2}}} u_n(x) d\mu_\sigma.$$

We now establish $L^1(\mu_\sigma)$-convergence. From estimates (a) and (b) above we have

$$C_1 C_3 \le \frac{\int_{\mathcal{A}_{t,\theta_0,\frac{\delta}{2}}} u_N d\mu_\sigma}{\int_{\mathcal{A}^c_{t,\theta_0,\frac{\delta}{2}}} u_N d\mu_\sigma} \le C_2 C_4.$$

Estimate (c) then shows that for some $\alpha < \beta < 1$ we have $\int u_{N+1}(x) d\mu_\sigma \le \beta \int u_N(x) d\mu_\sigma$. Since μ is supported on X_B^+ we conclude that

$$\int |M^{(n+m)N}_{-(\sigma+it)R_\sigma} g| d\mu = \int |L^{(n+m)N}_{-(\sigma+it)r_\sigma} w| d\mu \le \beta^N$$

where $g = w \circ \pi$.

Using a straightforward approximation argument a similar estimate can be seen to remain valid for any $g \in C^\alpha(X_A^+)$. We refer the reader to [3] for details.

We now establish uniform contraction of $M_{-(\sigma+it)R_\sigma}$ We can use the quasi-compactness of $M_{-\sigma R_\sigma}$ to choose $D > 0$ and $0 < \rho < 1$ (independent of $\sigma_0 \le \sigma \le h$) such that

$$|M^n_{-\sigma R_\sigma} g - \int g d\mu_\sigma|_\infty \le D||g||_{\alpha,t}\rho^n, \quad \forall n \ge 0$$

We can use this to write

$$\begin{aligned}
&||M^{(n+m)2N}_{-(\sigma+it)R_\sigma} g||_\infty \\
&\leq ||M^{(n+m)N}_{-\sigma R_\sigma}\left(M^{(n+m)N}_{-(\sigma+it)R_\sigma} g\right)||_\infty \\
&\leq \int |M^{(m+n)N}_{-(\sigma+it)R_\sigma} g| d\mu_\sigma + D||M^{(m+n)N}_{-(\sigma+it)R_\sigma} g||_{\alpha,t}\rho^{(n+m)N} \\
&\leq \beta^N + D\left(C|t|||g||_\infty + ||h'||_\infty\right)\rho^{(n+m)N} \\
&\leq \left(\beta^N + 3DC|t|\right)\rho^{(n+m)N} \\
&\leq E\gamma^N = E\gamma^N||g||_\infty
\end{aligned}$$

with $E > 0$ and $\max(\beta, \rho^{n+m}) < \gamma < 1$ (and where $E > 0$ can be assumed to be independent of $|t|$ provided we allow that $N = O(\log|t|)$).

We now establish norm contraction of $M_{\sigma+it}$ in Case II. We can use Proposition 11 to write

$$\begin{aligned}
\frac{1}{|t|}|M^{2N(n+m)}_{-(\sigma+it)R_\sigma} g|_\alpha &\leq \left(C||M^{N(n+m)}_{-(\sigma+it)R_\sigma} g||_\infty + \theta^{N(n+m)}\frac{1}{|t|}||(M^{N(n+m)}_{-(\sigma+it)R_\sigma} g)'||_\infty\right) \\
&\leq CE\gamma^N||g||_\infty + \alpha^{N(n+m)}\left(C||g||_\infty + \frac{1}{|t|}||g'||_\infty\right) \\
&\leq F\tau^N||w||_{\alpha,t}
\end{aligned}$$

with $F > 0$ and $\max(\gamma, \theta^{n+m}) < \tau < 1$.

7. The error term in the orbital counting functions

In this final section we complete the proof of Theorem 2 by obtaining the appropriate error term for $N(T)$. To do this it is convenient to introduce auxiliary functions defined by $\chi(T) = \sum_{e^{hd(x,\gamma x)} \leq T} 1$ and $\chi_1(T) = \int_1^T N(x)dx$. It is a simple observation that $\chi_1(T) = \sum_{e^{hd(x,\gamma x)} \leq T}\left(T - e^{hd(x,gx)}\right)$.

It will prove convenient to slightly modify the definition of the Poincaré series and to work with the following function

$$\eta_0(s) = \sum_{g\in\Gamma} e^{-shd(x,gx)}.$$

Observe that $\eta(sh) = \eta_0(s)$, so that, in particular, η_0 has a simple pole at $s = 1$, with residue $C > 0$, say. Clearly similar conclusions to those of Proposition 10 for $\eta(s)$ also hold for $\eta_0(s)$. By a slight abuse of notation, we shall assume that $\eta_0(s)$ is analytic in the strip $c_0 < Re(s) \leq 1$, for some $c_0 < 1$, except for a simple pole at $s = 1$, with the bounds

$$|\eta(\sigma + it)| \leq D|t|^\alpha, \quad \text{for } |t| \geq 1$$

for some $D > 0$ and $0 < \alpha < 1$.

The link between the analytic properties of $\eta_0(s)$ and the asymptotic behaviour of $\chi_1(T)$ is given by the following proposition.

Proposition 12. *If $c_0 < c < 1$ then*

$$\chi_1(T) = \frac{CT^2}{2} + \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \eta_0(s)\frac{T^{s+1}}{s(s+1)}ds.$$

In particular, $\chi_1(T) = CT^2/2 + O(T^{c+1})$.

Proof. As in the proof of Proposition 6, we use the identity

$$\frac{1}{2\pi i}\int_{d-i\infty}^{d+i\infty} \frac{y^{s+1}}{s(s+1)}ds = \begin{cases} 0 \text{ if } 0 < y \leq 1 \\ y-1 \text{ if } y > 1 \end{cases} \tag{7.1}$$

for $d > 1$. By applying (7.1) term by term to the series $\eta_0(s)$ we obtain

$$\chi_1(T) = \frac{1}{2\pi i}\int_{d-i\infty}^{d+i\infty} \eta_0(s)\frac{T^{s+1}}{s(s+1)}ds.$$

Using the estimate $\eta_0(s) = O(|t|^\alpha)$ we see that the integral $\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \eta(s)\frac{T^{s+1}}{s(s+1)}ds$ exists. Furthermore,

$$\begin{aligned}
&\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \eta_0(s)\frac{T^{s+1}}{s(s+1)}ds - \frac{1}{2\pi i}\int_{d-i\infty}^{d+i\infty} \eta_0(s)\frac{T^{s+1}}{s(s+1)}ds \\
&= \lim_{R\to+\infty} \frac{1}{2\pi i}\int_{\Gamma_R} \eta_0(s)\frac{T^{s+1}}{s(s+1)}ds
\end{aligned}$$

where Γ_R is the contour consisting of line segments joining $d - iR$, $d + iR$, $c + iR$ and $c - iR$. The result now follows by applying the residue theorem to see that

$$\frac{1}{2\pi i}\int_{\Gamma_R} \eta_0(s)\frac{T^{s+1}}{s(s+1)}ds = \frac{CT^2}{2}$$

for all $R > 0$.

As in section 4, we can use elementary arguments to translate the estimate for $\chi_1(T)$ into one for $\chi(T)$ and hence complete the proof of Theorem 2.

Proposition 13. *We have the estimate* $\chi(T) = CT + O(T^{(c+1)/2})$.

Proof. We define a function of T by $\epsilon = \epsilon(T) := T^{(c+1)/2}$. Since $T \to \chi(T)$ is monotone increasing we see that

$$\frac{\chi_1(T+\epsilon) - \chi_1(T)}{\epsilon} = \frac{1}{\epsilon}\int_T^{T+\epsilon} \chi(t)dt \geq \chi(T).$$

Observe that by Proposition 12

$$\begin{aligned}
\frac{\chi_1(T+\epsilon) - \chi_1(T)}{\epsilon} &= \frac{(T+\epsilon)^2}{2} - \frac{T^2}{2} + O\left(T^{c+1}\right) \\
&= T + O\left(\epsilon, \frac{T^{c+1}}{\epsilon}\right) \\
&= T + O\left(T^{(c+1)/2}\right)
\end{aligned}$$

Thus $\chi(T) \leq T + O\left(T^{(c+1)/2}\right)$. A similar argument based on the inequality

$$\frac{\chi_1(T) - \chi_1(T-\epsilon)}{\epsilon} = \frac{1}{\epsilon}\int_{T-\epsilon}^{T} \chi(t)dt \leq \chi(T)$$

gives that $\chi(T) \geq T + O\left(T^{(c+1)/2}\right)$. This completes the proof.

It only remains to observe that, since $N(T) = \chi(e^{hT})$, Proposition 13 is equivalent to Theorem 2.

References

1. R. Bowen, *Symbolic dynamics for hyperbolic flows*, Amer. J. Math. **95** (1973), 429- 460.
2. N. Chernov, *Markov approximations and decay of correlations for Anosov flows*, Preprint.
3. D. Dolgopiat, *On decay of correlations in Anosov flows*, Preprint.
4. W. Ellison and F. Ellison, *Prime numbers*, Hermann, Paris, 1985.
5. J. Feldman and D. Ornstein, *Semirigidity of horocycle flows over compact surfaces of variable negative curvature.*, Ergod. Th. and Dynam. Sys. **7** (1987), 49-72.
6. M. Gromov, *Three remarks on geodesic dynamics and the fundamental group*, preprint.
7. H. Huber, *Uber eine neue Klasse automorpher Funktionen und ein Gitterpunktproblem in der hyperbolischen Ebene*, Comment. Math. Helv. **30** (1956), 20-62.
8. H. Huber, *Zur analytischen Theorie hyperbolischer Raum formen und Bewegungsgruppen, II*, Math. Ann. **142** (1961), 385-398.
9. A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, C.U.P., New York, 1995.
10. G. Margulis, *On some applications of ergodic theory to the study of manifolds of negative curvature*, Funct. Anal. Appl. **3** (1969), 89–90.
11. W. Parry and M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Asterisque **187-188** (1990), 1-268.
12. M. Pollicott and R. Sharp, *The circle problem on surfaces of variable negative curvature*, Monat. Math. **123** (1997), 61-70.
13. M. Pollicott and R. Sharp, *Comparison theorems and orbit counting for hyperbolic geometry*, To appear in Trans. Amer. Math. Soc.
14. M. Pollicott and R. Sharp, *Poincaré series and comparison theorems for variable negative curvature*, Preprint.
15. M. Ratner, *Markov Partitions for Anosov flows on n-dimensional manifolds*, Israel J. Math. **15** (1973), 92-114.
16. D. Ruelle, *Zeta functions for for expanding maps and Anosov flows*, Invent. Math. **34** (1978), 231-242.
17. D. Ruelle, *Thermodynamic Formalism*, Addison-Wesley, New York, 1978.
18. D. Ruelle, *An extension of the theory of Fredholm determinants*, Publ. Math. IHES **72** (1990), 175-193.
19. A. Selberg, *Harmonic analysis and discontinuous subgroups in weakly symmetric Riemannian spaces with applications to Dirichlet series.*, J. Indian. Math. Soc. **20** (1956), 47-87.
20. C. Series, *Symbolic dynamics for geodesic flows*, Acta Math. **146** (1981), 103-128.
21. C. Series, *Geometrical Markov coding of geodesics on surfaces of constant negative curvature*, Ergod. Th. and Dynam. Sys. **6** (1986), 601-615.
22. E. Titchmarsh, *The theory of functions*, O.U.P., Oxford, 1939.

Department of Mathematics, University of Manchester, Oxford Road, Manchester M13 9PL, ENGLAND