Acta Math., 163 (1989), 1–55

Renewal theorems in symbolic dynamics, with applications to geodesic flows, noneuclidean tessellations and their fractal limits

by

STEVEN P. LALLEY

Purdue University
West Lafayette, IN, U.S.A.

Contents

0. Overview

This paper outlines a new approach to the asymptotic analysis of certain counting functions arising in the geometry of discrete groups. The approach is based on an analogue of the renewal theorem ([3], Chapter XI) for counting measures in symbolic dynamics.

The counting problems considered in this paper are mostly tied up with the ergodic behavior of the action of a discrete group at ∞. Some of these problems may be solved by other methods of noncommutative harmonic analysis, e.g., the Selberg trace formula, and in these cases the alternative methods may give sharper results (especially error estimates). Also, the methods developed here are not well suited for groups with parabolic elements, because of difficulties with the symbolic dynamics. However, our approach is suitable for certain problems that are apparently outside the scope of noncommutative harmonic analysis, in particular, problems directly concerned with the geometry of the limit set.

Let Γ be a Schottky group (Section 9; also [15]) acting on the hyperbolic plane H^2. (For simplicity, we shall state our main results for Schottky groups. These results hold for "most" finitely generated Fuchsian groups without parabolic elements; cf. Sections 10–13.) Consider the Riemann surface H^2/Γ; define $N(a)$ to be the number of closed geodesics on H^2/Γ with lengths $\leqslant a$. Let Λ be the limit set of Γ, and let δ be the Hausdorff dimension of Λ.

COROLLARY 11.2. *As* $a \to \infty$,

$$N(a) \sim \frac{e^{a\delta}}{a\delta}.$$

The corresponding result for cocompact Fuchsian groups ($\delta=1$) is a well known consequence of the Selberg trace formula ([5], Chapter 2). There are analogous results for closed geodesics on compact Riemannian manifolds of variable negative curvature [17] (but no hint of a proof appears in this paper) and, more generally, for periodic orbits of Axiom A flows [20], [19], [11].

Now let Γ be a Schottky group, as before, but consider Γ as a group of transformations of the plane $\mathbf{C}$ with the Euclidean metric d_E. Let z be a point of discontinuity for Γ, i.e., $z \in \mathbf{C} - \Lambda$.

THEOREM 10. *As* $\varepsilon \to 0$

$$\#\{\gamma \in \Gamma : d_E(\gamma z, \Lambda) > \varepsilon\} \sim C\varepsilon^{-\delta}$$

for a suitable constant $0 < C < \infty$. *Moreover, the uniform probability measure on* $\{\gamma z : d_E(\gamma z, \Lambda) > \varepsilon\}$ *converges weakly as* $\varepsilon \to 0$ *to the normalized* δ*-dimensional Hausdorff measure on* Λ.

This result is closely related to counting problems concerning noneuclidean lattice points ([6], [14], [17], [21], also Section 12), which are traditionally approached by spectral analysis of the Laplacean on H^2/Γ. It is not clear whether such methods can be adapted to obtain results like Theorem 10. The result concerning the asymptotic behavior of uniform distributions sheds some light on the construction of conformally invariant densities on Λ ([22], [31]).

Next, let $\mathcal{R}$ be the natural fundamental region for Γ (Section 9). The images $\gamma\mathcal{R}$, $\gamma\in\Gamma$, form a tessellation of $\mathbf{C}-\Lambda$ in which the tiles $\gamma\mathcal{R}$ accumulate at points of Λ. The (Euclidean) areas of these tiles converge to 0 as $\gamma\to\infty$.

THEOREM 11. *As* $\varepsilon\to 0$

$$\#\{\gamma\in\Gamma: \mathrm{Area}(\gamma\mathcal{R})>\varepsilon\}\sim C\varepsilon^{-\delta/2}$$

for a suitable constant $0<C<\infty$.

Consider now the limit set Λ with the induced Euclidean metric; keep in mind that Λ is compact. Let $N(\varepsilon)$ be the minimum number of ε-balls needed to cover Λ, and let $M(\varepsilon)$ be the maximum cardinality of an ε-separated subset of Λ (i.e., a subset F such that $d(x,y)>\varepsilon$ for all $x,y\in F$). Let G_ε be a subset of $\mathbf{C}$ of minimum cardinality such that every point of Λ is within distance ε of some point of G_ε; define P_ε to be the uniform probability measure on G_ε.

THEOREMS 12–13. *As* $\varepsilon\to 0$,

$$N(\varepsilon)\sim C\varepsilon^{-\delta},$$
$$M(\varepsilon)\sim C'\varepsilon^{-\delta}$$

for suitable constants $0<C,C'<\infty$. *Moreover, as* $\varepsilon\to 0$, P_ε *converges weakly to the normalized* δ*-dimensional Hausdorff measure on* Λ.

The counting functions $N(\varepsilon)$ and $M(\varepsilon)$ are used to define dimensional characteristics of Λ called the *metric entropy* and *capacity* [9]:

$$\text{metric entropy}=\lim_{\varepsilon\to 0}\frac{\log N(\varepsilon)}{\log\varepsilon^{-1}},$$
$$\text{capacity}=\lim_{\varepsilon\to 0}\frac{\log M(\varepsilon)}{\log\varepsilon^{-1}}.$$

Theorem 12 shows that for limit sets of Schottky groups the metric entropy, capacity, and Hausdorff dimension coincide. The asymptotic behavior of P_ε suggests that the δ-

dimensional Hausdorff measure be interpreted as a uniform distribution on Λ. Theorem 12, which is the hardest result of this paper, was suggested by an analogous but easier result for self-similar fractals [13]. For discussion of other aspects of the geometry of Λ, see [16].

The results mentioned thus far are all derived from an abstract renewal-type theorem in symbolic dynamics. Let (Σ, σ) be a one-sided shift of finite type (Section 1), e.g., $\Sigma=\Pi_1^\infty\{1,2,\ldots,l\}$ with the topology of coordinatewise convergence and $(\sigma x)_n=x_{n+1}$. For a continuous function $f: \Sigma\to\mathbf{C}$ define $S_nf=f+f\circ\sigma+\ldots+f\circ\sigma^{n-1}$. Let $f, g: \Sigma\to\mathbf{R}$ be Hölder continuous functions such that $g\geqslant 0$ but not identically zero and $S_nf>0$ on Σ for some $n\geqslant 1$. Define $N(a,x)$ for $a\in\mathbf{R}$, $x\in\Sigma$ by

$$N(a,x)=\sum_{n=0}^{\infty}\sum_{y:\,\sigma^n y=x} g(y)\,1\{S_nf(y)\leqslant a\}.$$

THEOREM 1. *If f is nonlattice then as $a\to\infty$,*

$$N(a,x)\sim C(x)e^{a\delta}$$

uniformly for $x\in\Sigma$, for suitable constants $0<\delta<\infty$, $0<C(x)<\infty$.

The hypothesis that f be nonlattice means that no function cohomologous to f (Section 1) maps Σ into a proper closed (additive) subgroup of $\mathbf{R}$. Explicit characterizations of the constants δ and $C(x)$ will be given in Section 2; δ depends only on f, while $C(x)$ depends linearly on g.

In applications it is often difficult to show that the relevant function f is nonlattice. For the applications given in this paper the nonlattice character of f derives from the fact that the geodesic flow on the unit tangent bundle TH^2/Γ is mixing relative to a certain invariant measure [24]. We shall use this to prove

COROLLARY 11.4. *The distortion cocycle of any discrete group containing a Schottky subgroup is nonlattice.*

See Section 11 for the definition of the distortion cocycle. Corollary 11.4 extends Theorem 6 of [32] (see also [30]).

Acknowledgement. The author has benefitted from many enlightening discussions with W. Lalley.

Part I. Renewal theorems in symbolic dynamics

1. Background: Shifts, suspension flows, thermodynamic formalism

A *shift of finite type* is defined as follows. Let A be an irreducible, aperiodic, $l\times l$ matrix of zeros and ones ($l>1$), called the *transition matrix*. Define Σ to be the space of all sequences taking values in the alphabet $\{1,2,\dots,l\}$ with transitions allowed by A, i.e.,

$$\Sigma=\left\{x\in\prod_{n=0}^{\infty}\{1,2,\dots,l\}:A(x_n,x_{n+1}\}=1,\forall n\}\right\}. \tag{1.1}$$

The space Σ is compact and metrizable in the product topology. Define the (forward) shift $\sigma:\Sigma\to\Sigma$ by $(\sigma x)_n=x_{n+1}$ for $n\geq 0$; observe that σ is continuous and surjective, but not 1–1. The system (Σ,σ) is topologically mixing ([1], Lemma 1.3).

Let $C(\Sigma)$ be the space of continuous, complex-valued functions on Σ. For $f\in C(\Sigma)$ and $0<\varrho<1$ define

$$\operatorname{var}_n(f)=\sup\{|f(x)-f(y)|:x_j=y_j,\forall 0\leq j\leq n\},$$

$$|f|_\varrho=\sup_{n\geq 0}\operatorname{var}_n(f)/\varrho^n,\quad\text{and}$$

$$\mathcal{F}_\varrho=\{f\in C(\Sigma):|f|_\varrho<\infty\}.$$

Elements of $\mathcal{F}_\varrho$ are called *Hölder continuous* functions. The space $\mathcal{F}_\varrho$, when endowed with the norm $\|\cdot\|_\varrho=|\cdot|_\varrho+\|\cdot\|_\infty$, is a Banach space.

For $f,g\in C(\Sigma)$ define $\mathcal{L}_fg\in C(\Sigma)$ by

$$\mathcal{L}_fg(x)=\sum_{y:\,\sigma y=x}e^{f(y)}g(y).$$

For each $\varrho\in(0,1)$ and $f\in\mathcal{F}_\varrho$, $\mathcal{L}_f:\mathcal{F}_\varrho\to\mathcal{F}_\varrho$ is a continuous linear operator; if f is real-valued then $\mathcal{L}_f$ is positive.

THEOREM A (Ruelle [25]). *For each real-valued $f\in\mathcal{F}_\varrho$, there is a simple eigenvalue $\lambda_f>0$ of $\mathcal{L}_f:\mathcal{F}_\varrho\to\mathcal{F}_\varrho$ with strictly positive eigenfunction h_f. The rest of the spectrum of $\mathcal{L}_f$ is contained in $\{z\in\mathbf{C}:|z|\leq\lambda_f-\varepsilon\}$ for some $\varepsilon>0$. There is a Borel probability measure ν_f on Σ such that $\mathcal{L}_f^*\nu_f=\lambda_f\nu_f$. If h_f is normalized so that $\int h_f d\nu_f=1$ then for every $g\in C(\Sigma)$*

$$\lim_{n\to\infty}\left\|\lambda_f^{-n}\mathcal{L}_f^ng-\left(\int g\,d\nu_f\right)h_f\right\|_\infty=0.$$

A proof may be found in [1], Chapter 1, also [25], [26]. Here $\mathcal{L}_f^*$ is the adjoint of $\mathcal{L}_f$.

The probability measure μ_f defined by $(d\mu_f/d\nu_f)=h_f$ is σ-invariant; it is called the *Gibbs measure* associated with f. For every Gibbs measure μ_f the dynamical system (Σ, μ_f, σ) is mixing, hence ergodic ([1], Proposition 1.14).

Functions $f, g \in C(\Sigma)$ are *cohomologous* if there exists $\varphi \in C(\Sigma)$ such that $f-g=\varphi-\varphi\circ\sigma$. If $f, g \in \mathcal{F}_\varrho$ are real-valued then $\mu_f=\mu_g$ iff f,g are cohomologous; if this is the case then $f-g=\varphi-\varphi\circ\sigma$ for some $\varphi \in \mathcal{F}_\varrho$ ([1], Theorem 1.28). Otherwise μ_f, μ_g are mutually singular. Observe that if $f-g=\varphi-\varphi\circ\sigma$ then

$$\mathcal{L}_f\psi = e^{-\varphi}\mathcal{L}_g(e^{\varphi}\psi);$$

consequently, for real-valued f, g, φ

$$\lambda_f = \lambda_g, \quad h_f = e^{-\varphi}h_g, \quad \nu_f = e^{\varphi}\nu_g.$$

For $f \in C(\Sigma)$ define $S_nf=f+f\circ\sigma+\ldots+f\circ\sigma^{n-1}$, $n\geq 1$, and $S_0f\equiv 0$. Functions $f, g \in \mathcal{F}_\varrho$ are cohomologous iff for every $n\geq 1$ and $x \in \Sigma$ such that $\sigma^n x=x$, $S_nf(x)=S_n g(x)$ ([1], Theorem 1.28). For real-valued $f \in \mathcal{F}_\varrho$, the Gibbs measure μ_f is the unique σ-invariant probability measure on Σ for which there exist constants $0<C_1\leq C_2<\infty$ such that

$$C_1 \leq \frac{\mu_f\{y \in \Sigma : y_i = x_i, \forall i = 0, 1, \ldots, n-1\}}{\lambda_f^{-n}\exp\{S_nf(x)\}} \leq C_2$$

for all $x \in \Sigma$ and $n\geq 0$.

The functional $P(f)=\log\lambda_f$ (the *pressure*) determines the moments of various functions relative to the Gibbs measures. In particular, for real-valued $f, g \in \mathcal{F}_\varrho$, $P(zf+g)$ is real-analytic in z and

$$\frac{dP(zf+g)}{dz} = \int_\Sigma f\, d\mu_{zf+g}, \quad z \in \mathbf{R} \tag{1.2}$$

([26], Theorem 5.26 and Exercise 5.5).

Define $\tilde{\Sigma}$ to be the space of all *double-ended* sequences with transitions allowed by A, i.e.,

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

and let $\sigma:\tilde{\Sigma}\to\tilde{\Sigma}$ be the forward shift. For a continuous function $f:\tilde{\Sigma}\to\mathbf{C}$ define $\operatorname{var}_n(f)=\sup\{|f(x)-f(y)|: x_i=y_i, \forall |i|\leq n\}$, and define $|f|_\varrho$, $\|f\|_\varrho$, $\tilde{\mathscr{F}}_\varrho$, S_nf, etc., as before. Notice that $\mathscr{F}_\varrho$ may be regarded as a closed, linear subspace of $\tilde{\mathscr{F}}_\varrho$ (namely, those functions in $\tilde{\mathscr{F}}_\varrho$ that depend only on the "forward" coordinates $x_0, x_1, \ldots$). Every $\tilde{f}\in\tilde{\mathscr{F}}_\varrho$ is cohomologous to a function $f\in\mathscr{F}_{\sqrt{\varrho}}$ ([1], Lemma 1.6). Observe that for every $f\in\mathscr{F}_\varrho$ the Gibbs measure μ_f extends naturally to a σ-invariant probability measure on $\tilde{\Sigma}$, again denoted by μ_f.

Let $f\in\tilde{\mathscr{F}}_\varrho$ be a strictly positive function on $\tilde{\Sigma}$. The *suspension flow* over the shift $(\tilde{\Sigma},\sigma)$ with *height function* f is defined as follows. Let $\tilde{\Sigma}_f=\{(x,t): x\in\tilde{\Sigma}, 0\leq t\leq f(x)\}$ with $(x,f(x))$ identified with $(\sigma x, 0)$. The flow T_s takes place on $\tilde{\Sigma}_f$; it is defined by

$$T_s(x,t)=(x,t+s) \quad \text{for} \quad t+s\leq f(x),\ s\geq 0,$$

$$T_{s_1}\circ T_{s_2}=T_{s_1+s_2} \quad \text{for} \quad s_1,s_2\geq 0.$$

See Figure 1.

Finally, we remark that the notion of a shift of finite type may be generalized in a useful but trivial way as follows. Instead of having the transition rules for sequences x in Σ (or $\tilde{\Sigma}$) depend only on adjacent entries x_n, x_{n+1}, one may have them depend on $x_{n-k}, x_{n-k+1}, \ldots, x_{n+k}$ for a fixed $k<\infty$. Such transition rules may be reduced to (1.1) by changing the alphabet $\{1,2,\ldots,l\}$ to a finite set of "words" from $\{1,2,\ldots,l\}$ of length $2k+1$. (This generalization will be needed for the symbolic dynamics in Section 10.)

2. Renewal measures and renewal theorems

Let $f\in\mathscr{F}_\varrho$ be a real-valued function such that for some n, S_nf is strictly positive on Σ, and let $g\in\mathscr{F}_\varrho$ be nonnegative but not identically zero. Define $N(a,x)$ for $a\in\mathbf{R}$, $x\in\Sigma$ by

$$N(a,x)=\sum_{n=0}^{\infty}\sum_{y:\sigma^n y=x} g(y)\,1\{S_nf(y)\leq a\}. \tag{2.1}$$

This is a distant cousin of the distribution function of the *renewal measure* in probability theory ([3], Chapter XI). Because S_nf is strictly positive for some n, $N(a,x)$ is finite for all $(a,x)\in\mathbf{R}\times\Sigma$; in fact $N(a,x)=0$ for $a\leq 0$, and $N(a,x)\leq C_1e^{C_2a}$ for suitable constants C_1, C_2 (see Lemma 8.1 below). Observe that $N(a,x)$ is nonnegative and nondecreasing in a. Most importantly, $N(a,x)$ satisfies an analogue of the *renewal equation*:

$$N(a,x)=\sum_{y:\sigma y=x} N(a-f(y),y)+g(x)\,1\{a\geq 0\}. \tag{2.2}$$

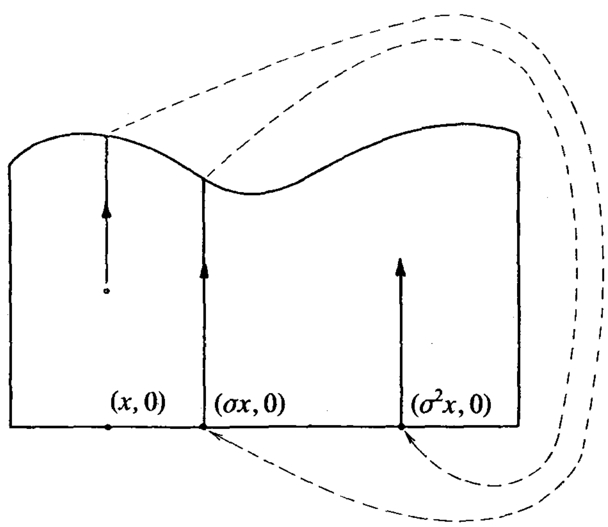

Fig. 1

The asymptotic comportment (as $a \to \infty$) of $N(a, x)$ depends on whether f is a *lattice* or *nonlattice* function. Say that f is a lattice function if f is cohomologous to a function taking values in a discrete subgroup of $\mathbf{R}$; otherwise, say that f is a nonlattice function.

PROPOSITION 2.1. *Let $f \in \mathcal{F}_\varrho$ be a real-valued function such that for some $n \geq 1$ the function $S_n f$ is strictly positive on Σ. Then $z \to \lambda_{zf}$ is strictly increasing for $z \in \mathbf{R}$, and there is a unique $\delta > 0$ such that*

$$\lambda_{-\delta f} = 1. \tag{2.3}$$

Henceforth, δ will be the unique real number such that (2.3) holds. The proof of Proposition 2.1 will be given at the end of the section.

THEOREM 1. *Assume that f is nonlattice. Then*

$$N(a, x) \sim C(x)\, e^{a\delta} \tag{2.4}$$

as $a \to \infty$, uniformly for $x \in \Sigma$.

From the renewal equation (2.2) one sees that if (2.4) holds then

$$C(x) = \sum_{\sigma y = x} e^{-\delta f(y)} C(y),$$

so $C(x)$ is an eigenfunction of the Perron-Frobenius operator $\mathscr{L}_{-\delta f}$. In fact the proof of Theorem 1 will show that

$$C(x)=\left(\frac{\int g\,d\nu_{-\delta f}}{\delta\int f\,d\mu_{-\delta f}}\right)h_{-\delta f}(x). \tag{2.5}$$

The lattice case is more complicated. If f takes values in a discrete additive subgroup G of $\mathbf{R}$, then clearly $N(a,x)$ is a step function in a with discontinuities only at elements of G. Thus, in general one would expect $N(a,x)$ to exhibit asymptotic periodicity in a.

THEOREM 2. *Assume that f is integer-valued, but not cohomologous to any function taking its values in a proper subgroup of the integers. Then*

$$N(a,x)\sim C(x)e^{[a]\delta} \tag{2.6}$$

as $a\to\infty$, uniformly for $x\in\Sigma$.

Here $[\cdot]$ denotes greatest integer. Once again $C(x)$ is an eigenfunction of $\mathscr{L}_{-\delta f}$; this time

$$C(x)=(1-e^{-\delta})^{-1}\left(\frac{\int g\,d\nu_{-\delta f}}{\int f\,d\mu_{-\delta f}}\right)h_{-\delta f}(x). \tag{2.7}$$

When f is not integer-valued but is cohomologous to an integer-valued function one again expects $e^{-a\delta}N(a,x)$ to be asymptotically periodic.

THEOREM 3. *Assume that f is cohomologous to an integer-valued function but not to any function taking values in a proper subgroup of the integers. Then there exists a bounded function $C(\beta,x)$, $\beta\in[0,1)$, $x\in\Sigma$, such that*

$$N(a,x)\sim C(a-[a],x)e^{a\delta} \tag{2.8}$$

as $a\to\infty$, uniformly for $x\in\Sigma$.

There is an explicit formula for $C(\beta,x)$ but it is neither simple nor illuminating, so we shall omit it.

Theorems 1–3 are comparable to theorems giving exponential decay of solutions to renewal equations ([3], Section XI.6, Theorem 2), and exponential growth and decay of solutions to Wiener-Hopf equations ([10], Theorem 15.4). The proofs of Theorems 1–3 will be given in Sections 7–8.

Proof of Proposition 2.1. Since the Gibbs measure μ_{zf} is σ-invariant for each $z \in \mathbf{R}$,

$$\int_\Sigma (S_n f/n)\, d\mu_{zf} = \int_\Sigma f\, d\mu_{zf} = (d/dz)(\log \lambda_{zf}),$$

by (1.2). Since $S_n f/n \geq \varepsilon > 0$ for some n, it follows that $\log \lambda_{zf}$ is strictly increasing in z; moreover, $(d/dz)\log \lambda_{zf} \geq \varepsilon$ so $\lambda_{zf} \to 0$ as $z \to -\infty$ and $\lambda_{zf} \to \infty$ as $z \to \infty$. Now λ_{zf} is analytic in z (cf. [26], Theorem 5.26) so $\lambda_{zf} = 1$ has a unique solution $z \in \mathbf{R}$.

Recall that the transition matrix A is irreducible and aperiodic. Consequently,

$$\mathscr{L}_0^n 1(x) = \sum_{i=1}^{l} A^n(i, x_0) \to \infty$$

as $n \to \infty$, for each $x \in \Sigma$ (here 1 denotes the function in $\mathscr{F}_\varrho$ that is identically 1). It follows from Theorem A that $\lambda_0 > 1$. Therefore, since λ_{zf} is increasing, if $\lambda_{zf} = 1$ then $z < 0$. □

3. A modification for finite sequences

Define Σ_* to be the set of all *finite* sequences from the alphabet $\{1, 2, \ldots, l\}$ with transitions allowed by A, i.e.,

$$\Sigma_* = \{\xi\} \cup \Big(\bigcup_{n=0}^{\infty} \{(x_0, x_1, \ldots, x_n) : A(x_i, x_{i+1}) = 1\}\Big)$$

(ξ is the empty sequence). The forward shift σ_* on Σ_* (or $\Sigma \cup \Sigma_*$) is defined in the obvious way, in particular,

$$\sigma_*(x_0, x_1, \ldots, x_n) = (x_1, x_2, \ldots, x_n),$$
$$\sigma_*(\xi) = \xi.$$

Sequences in Σ_* may be extended to infinite sequences by adjoining an additional symbol 0 to the alphabet $\{1, 2, \ldots, l\}$ and making the correspondence

$$(x_0, x_1, \ldots, x_n) \leftrightarrow (x_0, x_1, \ldots, x_n, 0, 0, \ldots).$$

Thus the shift σ_* on $\Sigma \cup \Sigma_*$ is a shift of finite type. The transition matrix A is extended by $A(0,0) = 1, A(i, 0) = 1, A(0, i) = 0$ for $i = 1, 2, \ldots, l$. Unfortunately, the extension of A is not irreducible, so $(\Sigma \cup \Sigma_*, \sigma_*)$ is not topologically mixing and therefore the thermodynamic formalism of Section 1 does not apply directly. However, $\Sigma \cup \Sigma_*$ is compact and metrizable in the product topology, and $\mathrm{var}_n f$, $|f|_\varrho$, $\|f\|_\varrho$, $S_n f$, cohomology, etc., may be defined as in Section 1. Define $\mathscr{F}_\varrho(\Sigma\cup\Sigma_*)=\{f\in C(\Sigma\cup\Sigma_*): |f|_\varrho<\infty\}$.

Let $f_*\in\mathscr{F}_\varrho(\Sigma\cup\Sigma_*)$ be a fixed, real-valued function, and let f be its restriction to Σ; assume that for some n, $S_n f$ is strictly positive. Let g_* be a *nonnegative* function on $\Sigma\cup\Sigma_*$, not identically zero, such that $\mathrm{var}_m g_*=0$ for some $m<\infty$, and let g be the restriction of g_* to Σ. For $a\in\mathbf{R}$, $x\in\Sigma\cup\Sigma_*$ define

$$N_*(a,x)=\sum_{n=0}^{\infty}\sum_{\substack{y:\sigma_*^n y=x\\ y\neq\xi}} g_*(y)\,1\{S_n f_*(y)\leqslant a\};\tag{3.1}$$

as in Section 2, $N_*(a,x)$ is finite, identically zero for $a\leqslant 0$, and grows at most exponentially as $a\to\infty$. There is again a renewal equation:

$$N_*(a,x)=\sum_{\substack{y:\sigma_* y=x\\ y\neq\xi}} N_*(a-f_*(y),y)+g_*(x)\,1\{a\geqslant 0\}.\tag{3.2}$$

THEOREM 4. *Let $\delta>0$ be the unique real such that $\lambda_{-\delta f}=1$. If f is nonlattice then*

$$N_*(a,x)\sim C_*(x)\,e^{a\delta}\tag{3.3}$$

as $a\to\infty$, uniformly for $x\in\Sigma\cup\Sigma_*$, where

$$C_*(x)=\left(\frac{\int g\,d\nu_{-\delta f}}{\delta\int f\,d\mu_{-\delta f}}\right)h_*(x)\tag{3.4}$$

and $h_*(x)$ is the unique positive continuous function on $\Sigma\cup\Sigma_*$, satisfying

$$h_*(x)=h_{-\delta f}(x),\quad x\in\Sigma,\tag{3.5}$$

$$h_*(x)=\sum_{\substack{y:\sigma_* y=x\\ y\neq\xi}} e^{-\delta f_*(y)}h_*(y),\quad x\in\Sigma_*.\tag{3.6}$$

There are analogous results for the lattice cases, which we shall refrain from stating. The proof of Theorem 4 will be given in Section 6.

Theorem 4 differs from Theorem 1 in that more restrictive hypotheses are imposed on $g(x)$. These may be relaxed somewhat; however, for the applications given in this paper this is unnecessary.

For many counting problems in symbolic dynamics Theorems 1–4 apply in a very direct fashion to give the asymptotic behavior of the relevant counting function.

Examples of such problems will be encountered in Sections 5 and 12. There are more subtle counting problems, however, in which the counting function cannot be written in the form (2.1) or (3.1), but to which Theorems 1–4 may be applied indirectly. An example will be given in Section 13. In this example, and certain others like it, one must deal with a function of the form

$$N_*(a,x;t)=\sum_{n=0}^{\infty}\sum_{y:\sigma_*^n y=x} g_*(y)\,1\{a\leq S_{n+1}f_*(y)\leq a+t;S_jf_*(y)\leq a,\forall j\leq n\} \tag{3.7}$$

where $a,t\geq 0$, $x\in\Sigma_*\cup\Sigma$, and $g_*\geq 0$ is a function such that $\mathrm{var}_k g_*=0$ for some $k<\infty$.

COROLLARY 3.1. *Assume that f_* satisfies the hypotheses of Theorem 4 and assume that $f_*(x)>0$ for all $x\in\Sigma\cup\Sigma_*$, $x\neq\xi$. Then as $a\to\infty$*

$$N_*(a,x;t)\sim F(x,t)\,e^{a\delta}, \tag{3.8}$$

uniformly for $x\in\Sigma\cup\Sigma_*$ and t in any compact subset of $(0,\infty)$, where

$$F(x,t)=C_*(x)\,(e^{\delta((t-f_*(x))\wedge 0)}-e^{-\delta f_*(x)}) \tag{3.9}$$

and $C_*(x)$ is defined by (3.4).

Note. $a\wedge b$ is the minimum of a and b.

Proof. It is easily verified that

$$a<S_{n+1}f_*(y)\leq a+t,$$
$$S_jf_*(y)\leq a,\quad \forall j\leq n,$$
$$\sigma_*^n y=x$$

occurs iff

$$a-f_*(x)<S_nf_*(y)\leq a+((t-f_*(x))\wedge 0),$$
$$\sigma_*^n y=x.$$

Hence

$$N_*(a,x;t)=N_*(a+((t-f_*(x))\wedge 0),x)-N(a-f_*(x),x)$$

where $N_*(a,x)$ is defined by (3.1), so (3.8) follows from (3.3). □

A similar result holds when the hypothesis $f_*>0$ is dropped.

Now let $G(t)$, $t\in\mathbf{R}$ be a nonnegative, monotone function (either nondecreasing or nonincreasing); let $g_*(x)$ be as before. For $x\in\Sigma_*\cup\Sigma$ define

$$N_G(a,x)=\sum_{n=0}^{\infty}\sum_{y:\sigma_*^n y=x} g_*(y)\,G(S_{n+1}f_*(y)-a)\,1\{a<S_{n+1}f_*(y);\,S_jf_*(y)\le a,\ \forall j\le n\}. \tag{3.10}$$

COROLLARY 3.2. *Under the hypotheses of Corollary* 3.1, *for each* $x\in\Sigma_*\cup\Sigma$, $x\neq\xi$,

$$N_G(a,x)\sim e^{a\delta}\int_0^\infty G(t)\,F(x,dt) \tag{3.11}$$

as $a\to\infty$, *uniformly for* $x\in\Sigma_*\cup\Sigma$.

This follows from Corollary 3.1 by a routine argument. Note that the measure $F(x,dt)$ is supported by the interval $[0,\|f_*\|_\infty]$ so the integral $\int_0^\infty$ could be changed to $\int_0^{\|f_*\|_\infty}$.

An important consequence of Corollary 3.1 is that $F(x,t)$ is jointly continuous in x,t. This implies

COROLLARY 3.3. *For every* $\varepsilon>0$ *there exists* $\alpha_*>0$ *such that if* $G(t)$ *is any monotone function satisfying* $|G(C+1)-G(0)|\le 1$ *where* $C=\|f_*\|_\infty$ *and if* $0<\alpha<\alpha_*$ *then*

$$\left|\int_0^\infty G(t+\alpha)\,F(x,dt)-\int_0^\infty G(t)\,F(x,dt)\right|<\varepsilon \tag{3.12}$$

for every $x\in\Sigma\cup\Sigma_*$, $x\neq\xi$.

Proof. Since $F(x,t)=F(x,C)$ for $t\ge C$, the limits of integration may be changed from $[0,\infty)$ to $[0,C]$. Fubini's theorem (integration by parts) implies that the difference between the two integrals is

$$-\int_{[0,\alpha]}F(x,t)\,G(dt)+\int_{(\alpha,C]}(F(x,t-\alpha)-F(x,t))\,G(dt)+\int_{(C,C+\alpha]}F(x,t-\alpha)\,G(dt).$$

The result now follows easily from the uniform continuity and boundedness of $F(x,t)$. □

4. Equidistribution theorems

Once again let $f\in\mathscr{F}_\varrho$ be a real-valued function such that S_nf is strictly positive on Σ for some $n\ge1$, and let $x\in\Sigma$ be a fixed but arbitrary admissible sequence. In this section we shall discuss the distribution of those $y\in\Sigma$ such that $\sigma^n y=x$ and $S_nf(y)\le a$ for some n.

Let $P^{x,a}$ be the probability measure that attaches probability $1/N(a,x)$ to each element of

$$F^{x,a}=\{y\in\Sigma\colon \sigma^n y=x,\ S_nf(y)\leqslant a \text{ for some } n\geqslant 0\},$$

where

$$N(a,x)=\#(F^{x,a}).$$

THEOREM 5. *Assume that f is nonlattice. Then as $a\to\infty$*

$$P^{x,a}\xrightarrow{D}\nu_{-\delta f} \tag{4.1}$$

uniformly for $x\in\Sigma$.

Note. $\xrightarrow{D}$ (convergence in distribution) means that for each continuous $g\colon\Sigma\to\mathbf{R}$

$$\int_\Sigma g\,dP^{x,a}\to\int_\Sigma g\,d\nu_{-\delta f} \tag{4.2}$$

as $a\to\infty$.

Proof. Let $g\in\mathscr{F}_\varrho$ be nonnegative; then

$$\int_\Sigma g\,dP^{x,a}=\frac{\sum_{n=0}^{\infty}\sum_{y:\sigma^n y=x} g(y)\,1\{S_nf(y)\leqslant a\}}{\sum_{n=0}^{\infty}\sum_{y:\sigma^n y=x} 1\{S_nf(y)\leqslant a\}}\to\int_\Sigma g\,d\nu_{-\delta f}$$

as $a\to\infty$, uniformly in x, by Theorem 1 and (2.5). Since any nonnegative, continuous function may be uniformly approximated by functions in $\mathscr{F}_\varrho$, (4.2) and hence (4.1) follow. □

For $g\in C(\Sigma)$ define $\bar g=\int g\,d\mu_{-\delta f}$, and for $y\in F^{x,a}$ let $n(y)$ be the largest $n\geqslant 0$ such that $\sigma^n y=x$ and $S_n f(y)\leqslant a$. Unless x is periodic there is only one such n for any $y\in F^{x,a}$. If x is periodic the number of $y\in F^{x,n}$ such that $\sigma^n y=x$ has more than one solution $n\geqslant 0$ grows linearly in a, since $S_n f>0$ for some m.

THEOREM 6. *Assume that f is nonlattice. Then for each $g\in C(\Sigma)$ and each $\varepsilon>0$*

$$P^{x,a}\left\{y:\ \left|\frac{S_{n(y)}g(y)}{n(y)}-\bar{g}\right|>\varepsilon\right\}\to 0, \tag{4.3}$$

$$P^{x,a}\left\{y:\ \left|\frac{a}{n(y)}-\bar{f}\right|>\varepsilon\right\}\to 0 \tag{4.4}$$

as $a\to\infty$, uniformly for $x\in\Sigma$.

Proof. Let $z>\delta>0$, so $\lambda_{-zf}<1$ (Proposition 2.1). Then

$$\begin{aligned}\sum_{n=m}^{\infty}\sum_{y:\sigma^n y=x} 1\{S_nf(y)\leqslant m(\bar{f}-\varepsilon)\} &\leqslant \sum_{n=m}^{\infty}\sum_{y:\sigma^n y=x} e^{-zS_nf(y)}e^{zm(\bar{f}-\varepsilon)}\\ &=\sum_{n=m}^{\infty}(\mathscr{L}^n_{-zf}1)(x)\,e^{zm(\bar{f}-\varepsilon)}\\ &\sim \lambda^m_{-zf}(1-\lambda_{-zf})^{-1}C(x)\,e^{zm(\bar{f}-\varepsilon)}\end{aligned} \tag{4.5}$$

as $m\to\infty$ uniformly for $x\in\Sigma$, by Ruelle's theorem (Section 1). Theorem 1 implies that

$$\sum_{n=1}^{\infty}\sum_{y:\sigma^n y=x} 1\{S_nf(y)\leqslant m(\bar{f}-\varepsilon)\}\sim C'(x)\,e^{m(\bar{f}-\varepsilon)\delta} \tag{4.6}$$

as $m\to\infty$. Now $(d/dz)(\log\lambda_{-zf})_{z=\delta}=-\bar{f}$ (cf. (1.2)) so there exists $z>\delta$ such that

$$\lambda_{-\delta f}e^{(z-\delta)(\bar{f}-\varepsilon)}<1.$$

Therefore (4.5) and (4.6) imply that

$$\frac{\displaystyle\sum_{n=m}^{\infty}\sum_{\sigma^n y=x} 1\{S_nf(y)\leqslant m(\bar{f}-\varepsilon)\}}{\displaystyle\sum_{n=1}^{\infty}\sum_{\sigma^n y=x} 1\{S_nf(y)\leqslant m(\bar{f}-\varepsilon)\}}\to 0$$

as $m\to\infty$ uniformly for $x\in\Sigma$, which in turn implies that

$$P^{x,a}\{y:\ n(y)\bar{f}\geqslant a(1+\bar{\varepsilon})\}\to 0$$

as $a\to\infty$ uniformly for $x\in\Sigma$.

A similar argument, this time using $\delta>z>0$, implies that

$$\frac{\sum_{n=1}^{m} \sum_{\sigma^n y=x} 1\{S_n f(y) \leqslant m(\bar{f}+\varepsilon)\}}{\sum_{n=1}^{\infty} \sum_{\sigma^n y=x} 1\{S_n f(y) \leqslant m(\bar{f}+\varepsilon)\}} \to 0$$

as $m \to \infty$, and hence

$$P^{x,a}\{y\colon n(y)\bar{f} \leqslant a(1-\bar{\varepsilon})\} \to 0$$

as $a \to \infty$ uniformly for $x \in \Sigma$. This proves (4.4).

In proving (4.3) it suffices to consider *positive* $g \in \mathscr{F}_\varrho$. For such g,

$$\left(\frac{d}{dz} \log \lambda_{zg-\delta f}\right)_{z=0} = \bar{g} > 0,$$

so for all sufficiently small $z>0$, $\lambda_{-zg-\delta f} e^{z(\bar{g}-\varepsilon)} < 1$. Consequently,

$$\begin{aligned} &\sum_{n=0}^{\infty} \sum_{y:\sigma^n y=x} 1\{S_n g(y) \leqslant n(\bar{g}-\varepsilon)-a\varepsilon; S_n f(y) \leqslant a\} \\ &\quad \leqslant \sum_{n=0}^{\infty} \sum_{y:\sigma^n y=x} \exp\{-zS_n g(y)+zn(\bar{g}-\varepsilon)-\delta S_n f(y)+a(\delta-z\varepsilon)\} \\ &\quad = \sum_{n=0}^{\infty} (\mathscr{L}^n_{-zg-\delta f} 1)(x)\, e^{zn(\bar{g}-\varepsilon)} e^{a\delta-az\varepsilon} \\ &\quad \leqslant C e^{a\delta-az\varepsilon} \end{aligned}$$

for some $C<\infty$. Since

$$\sum_{n=0}^{\infty} \sum_{y:\sigma^n y=x} 1\{S_n f(y) \leqslant a\} \sim C(x)\, e^{a\delta}$$

by Theorem 1, it follows that

$$\frac{\sum_{n=0}^{\infty} \sum_{y:\,\sigma^n y=x} 1\{S_n g(y) \leqslant n(\bar{g}-\varepsilon)-a\varepsilon; S_n f(y) \leqslant a\}}{\sum_{n=0}^{\infty} \sum_{y:\,\sigma^n y=x} 1\{S_n f(y) \leqslant a\}} \to 0 \tag{4.7}$$

as $a \to \infty$, uniformly for $x \in \Sigma$. A similar argument shows that

$$\frac{\displaystyle\sum_{n=0}^{\infty}\sum_{y:\sigma^n y=x} 1\{S_n g(y)\geq n(\bar{g}+\varepsilon)+a\varepsilon;\, S_n f(y)\leq a\}}{\displaystyle\sum_{n=0}^{\infty}\sum_{y:\sigma^n y=x} 1\{S_n f(y)\leq a\}}\to 0 \tag{4.8}$$

as $a\to\infty$, uniformly in x. The result (4.3) now follows easily from (4.4), (4.7), and (4.8). □

Theorems 5 and 6 extend to the setting considered in Section 3. We will not need such extensions, however, so we shall not state them here.

5. Periodic orbits of suspension flows

Consider a suspension flow over the two-sided shift $(\bar{\Sigma}, \sigma)$ with height function g. Clearly, every $x\in\bar{\Sigma}$ such that $\sigma^n x=x$, $\sigma^i x\neq x$, $i=1,2,\ldots,n-1$ for some n lies on a periodic orbit τ of the flow; the period of τ is $S_n g(x)$. Since there are infinitely many periodic sequences $x\in\bar{\Sigma}$, it follows that there are infinitely many periodic orbits τ.

Recall that g is cohomologous to a function f that depends only on the "forward" coordinates $x_1, x_2, \ldots$; f may be considered a function on Σ. It is not difficult to show that $S_n f$ is strictly positive for some n, since $g>0$. The suspension flow is *topologically mixing* if f (equivalently, g) is nonlattice. Let $\delta>0$ be as in Proposition 2.1.

Define a measure M on $\bar{\Sigma}_g=\{(x,t)\colon x\in\bar{\Sigma},\ 0\leq t<g(x)\}$ by

$$M(A\times B)=\mu_{-\delta f}(A)\,m(B)\Big/\int f\,d\mu_{-\delta f} \tag{5.1}$$

for any rectangle $A\times B\subset\bar{\Sigma}_g$, where m is Lebesgue measure. Since $\int f\,d\mu_{-\delta f}=\int g\,d\mu_{-\delta f}$, M is a probability measure. It is easily verified that M is invariant for the flow.

Denote periodic orbits of the suspension flow by τ; let $\lambda(\tau)$ be the (minimal) period of τ. For any continuous function $G\colon\bar{\Sigma}_g\to\mathbf{R}$ let $\tau(G)$ be the mean value of G on τ, i.e., the integral of G over τ divided by $\lambda(\tau)$.

THEOREM 7. *Assume that the suspension flow is topologically mixing. Then for any* $\varepsilon>0$,

$$\#\{\tau\colon \lambda(\tau)\leq a\}\sim e^{a\delta}/a\delta, \tag{5.2}$$

$$\frac{\#\{\tau\colon \lambda(\tau)\leqslant a \text{ and } |\tau(G)-\int G\,dM|>\varepsilon\}}{\#\{\tau\colon \lambda(\tau)\leqslant a\}}\to 0 \tag{5.3}$$

as $a\to\infty$.

The first statement is equivalent to the main result of [20]; the second statement strengthens the main result of [19]. For more refined results, see [11].

Proof. Suppose that τ is a periodic orbit of the suspension flow that crosses the "floor" $\{(x,0)\colon x\in\bar{\Sigma}\}$ exactly n times. Then there is an n-periodic sequence x such that the crossings occur at $(x,0),(\sigma x,0),(\sigma^2 x,0),\ldots,(\sigma^{n-1}x,0)$, and

$$\lambda(\tau)=S_n g(x)=S_n f(x).$$

Consequently,

$$\#\{\tau\colon \lambda(\tau)\leqslant a\}=\sum_{n=1}^{\infty} n^{-1}\#\{x\in\Sigma\colon x \text{ has least period } n \text{ and } S_n f(x)\leqslant a\}.$$

If x is d-periodic, with $d|n$, and $S_n f(x)\leqslant a$ then $S_d f(x)\leqslant a/d\leqslant a/2$. Hence,

$$\begin{aligned}&\sum_{n=1}^{\infty} n^{-1}\#\{x\in\Sigma\colon x \text{ is } n\text{-periodic, } S_n f(x)\leqslant a\}\\&\qquad-\sum_{n=1}^{\infty} n^{-1}\#\{x\in\Sigma\colon x \text{ is } n\text{-periodic, } S_n f(x)\leqslant a/2\}\\&\qquad\leqslant \#\{\tau\colon \lambda(\tau)\leqslant a\}\\&\qquad\leqslant \sum_{n=1}^{\infty} n^{-1}\#\{x\in\Sigma\colon x \text{ is } n\text{-periodic, } S_n f(x)\leqslant a\}.\end{aligned}$$

Therefore, to prove (5.2) it suffices to prove

$$\sum_{n=1}^{\infty} n^{-1}\#\{x\in\Sigma\colon x \text{ is } n\text{-periodic, } S_n f(x)\leqslant a\}\sim\frac{e^{a\delta}}{a\delta}. \tag{5.4}$$

Let k be a fixed large integer; choose $x^{(1)},x^{(2)},\ldots,x^{(m)}\in\Sigma$ such that $x_1^{(i)}x_2^{(i)}\ldots x_k^{(i)}$, $i=1,2,\ldots,m$, are the distinct sequences of length k in Σ_*. Assume that each $x^{(i)}$ is an aperiodic sequence. Define $g_i\colon \Sigma\mapsto\mathbf{R}$ by

$$g_i(x)=1\{x_j=x_j^{(i)},\ \forall j=1,2,\ldots,k\};$$

observe that $\Sigma_{i=1}^m g_i\equiv 1$.

Consider $x\in\Sigma$ such that $\sigma^n x=x^{(i)}$ and $g_i(x)=1$. Let $\tilde{x}$ be the n-periodic sequence $\tilde{x}=x_1x_2\ldots x_nx_1x_2\ldots x_n\ldots$; then

$$|S_nf(\tilde{x})-S_nf(x)|\leqslant\sum_{j=k}^{\infty}\operatorname{var}_j(f)=\varepsilon_k.$$

Keep in mind that $\varepsilon_k\to 0$ as $k\to\infty$. Because of our choice of $x^{(1)},\ldots,x^{(m)}$, there is a 1-1 correspondence between n-periodic sequences and sequences x such that $\sigma^n x=x^{(i)}$ and $g_i(x)=1$ for some $i=1,2,\ldots,m$, provided $n\geqslant k$ (the correspondence is given by $x\leftrightarrow\tilde{x}$). Hence, for $n\geqslant k$,

$$\sum_{i=1}^{m}N_n^{(i)}(a-\varepsilon_k)\leqslant\#\{x\colon x\text{ is }n\text{−periodic, }S_nf(x)\leqslant a\}\leqslant\sum_{i=1}^{m}N_n^{(i)}(a+\varepsilon_k)\tag{5.5}$$

where

$$N_n^{(i)}(a)=\#\{x\colon\sigma^n x=x^{(i)},\ g_i(x)=1,\ S_nf(x)\leqslant a\}.$$

Now as $a\to\infty$, $\#\{\tau\colon\lambda(\tau)\leqslant a\}\to\infty$, so the contribution to the sum (5.4) from the terms $n<k$ is negligible as $a\to\infty$; consequently we may ignore the fact that (5.5) may not hold when $n<k$.

Recall that f satisfies the hypotheses of Theorem 1. Therefore, as $a\to\infty$

$$N^{(i)}(a)=\sum_{n=1}^{\infty}N_n^{(i)}(a)=\sum_{n=1}^{\infty}\sum_{x\colon\sigma^n x=x^{(i)}}g_i(x)\,1\{S_nf(x)\leqslant a\}\sim C_ie^{a\delta}\tag{5.6}$$

where

$$C_i=\left(\frac{\int g_i\,d\nu_{-\delta f}}{\delta\int f\,d\mu_{-\delta f}}\right)h_{-\delta f}(x^{(i)})>0.$$

Now Theorem 6 implies that for nearly all x such that $\sigma^n x=x^{(i)}$ and $S_nf(x)\leqslant a$ for some $n\geqslant 1$,

$$n\approx a\Big/\int f\,d\mu_{-\delta f}.$$

when a is large. Since for $\eta<(\|f\|_\infty)^{-1}$

$$\sum_{1\leqslant n\leqslant a\eta}N_n^{(i)}(a)\leqslant N^i(a\eta\|f\|_\infty)=o(e^{a\delta}/a),$$

it follows from (5.6) that as $a \to \infty$,

$$\sum_{n=1}^{\infty} n^{-1} N_n^{(i)}(a) \sim C_i \left(\int f \, d\mu_{-\delta f} \right) e^{a\delta} \Big/ a. \tag{5.7}$$

Observe that as $k \to \infty$, $\varepsilon_k \to 0$ and

$$\sum_{i=1}^{m} C_i \int f d\mu_{-\delta f} = \sum_{i=1}^{m} \delta^{-1} \left(\int g_i \, d\nu_{-\delta f} \right) h_{-\delta f}(x^{(i)}) \to \delta^{-1} \int h_{-\delta f} d\nu_{-\delta f} = \delta^{-1}.$$

Therefore, (5.4), (5.5), and (5.7) imply (5.2).

It remains to prove (5.3). Let G: $\bar{\Sigma}_g \to \mathbf{R}$ be continuous; define h: $\bar{\Sigma} \to \mathbf{R}$ by

$$h(x) = \int_0^{g(x)} G(x, t) \, dt.$$

Now h is continuous, so it may be uniformly approximated by $\tilde{h} \in \bar{\mathcal{F}}_\varrho$. Recall that any $\tilde{h} \in \bar{\mathcal{F}}_\varrho$ is cohomologous to some $\psi \in \mathcal{F}_\varrho$. Thus, if τ is a periodic orbit passing through $(x, 0)$, where x is an n-periodic sequence, then

$$\tau(G) \approx S_n \psi(x) / S_n f(x).$$

By the same reasoning as that leading up to (5.4) we conclude that to prove (5.3) it suffices to prove that

$$\sum_{n=1}^{\infty} n^{-1} \# \left\{ x \colon x \text{ is } n\text{-periodic}, \ S_n f(x) \le a, \ \left| S_n \psi - n \int \psi \, d\mu_{-\delta f} \right| > n\varepsilon \right\} = o(e^{a\delta}/a).$$

This may be accomplished by virtually the same argument used in proving (5.2), this time using (4.3) of Theorem 6. □

6. Proof of Theorem 4

Theorem 4 is a relatively straightforward consequence of Theorem 1. Observe that (3.3) holds for $x \in \Sigma$ by Theorem 1. Furthermore, if (3.3) holds for all $x \in \Sigma \cup \Sigma_*$ then the renewal equation (3.2) implies that

$$C(x) = \sum_{\substack{y : \sigma_* y = x \\ y \neq \xi}} e^{-\delta f_*(y)} C(y), \quad x \in \Sigma \cup \Sigma_*,$$

and (2.5) implies that

$$C(x)=\left(\frac{\int g\,d\nu_{-\delta f}}{\delta\int f\delta\mu_{-\delta f}}\right)h_{-\delta f}(x),\quad x\in\Sigma.$$

Therefore, to establish that (3.4) holds for all $x\in\Sigma\cup\Sigma_*$ it suffices to show that $C(x)$ is continuous and that the nonnegative solution to (3.5)–(3.6) is unique.

LEMMA 6.1. *There is at most one nonnegative, continuous $h_*(x)$ on $\Sigma\cup\Sigma_*$ satisfying* (3.5)–(3.6).

Proof. First notice that any solution $h_*(x)$ must be strictly positive on $\Sigma\cup\Sigma_*$. It is positive on Σ because of (2.5), since $h_{-\delta f}>0$. Hence it is positive near Σ, by continuity. But for any $x\in\Sigma_*$ there exists $y\in\Sigma_*$ near Σ such that $\sigma_*^n y=x$ for some n (since the matrix A is irreducible); iterating (3.6) n times shows that $h_*(x)\geq e^{-\delta S_n f_*(y)}h_*(y)>0$.

For $x,y\in\Sigma_*$, $y\neq\xi$, define

$$k(x,y)=\begin{cases}e^{-\delta f_*(y)}(h_*(y)/h_*(x)) & \text{if } \sigma_* y=x\\ 0 & \text{otherwise.}\end{cases}$$

Then for each $x\in\Sigma_*$, $\Sigma_{y:\sigma_* y=x}k(x,y)=1$. Define

$$k^n(x,y)=\sum_{(y_1,y_2,\dots,y_{n-1})}k(x,y_1)\,k(y_1,y_2)\dots k(y_{n-1},y);$$

then by induction $\Sigma_{y:\sigma_*^n y=x}k^n(x,y)=1$. Now suppose $h'_*(x)$ is another nonnegative, continuous solution to (3.5)–(3.6). Then

$$\frac{h'_*(x)}{h_*(x)}=\sum_y k(x,y)\left(\frac{h'_*(y)}{h_*(y)}\right)=\sum_y k^n(x,y)\left(\frac{h'_*(y)}{h_*(y)}\right).$$

As $n\to\infty$, any $y\in\Sigma_*$ such that $k(x,y)>0$ (i.e., $\sigma_*^n y=x$) gets increasingly close to Σ. But near Σ, $h'_*/h_*\approx 1$, by (3.5) and continuity. It follows that $h'_*(x)/h_*(x)=1$ for every $x\in\Sigma_*$. □

To extend (3.3) to $x\in\Sigma_*$ we will use the fact that $g_*(x)$ depends on only finitely many coordinates of x.

LEMMA 6.2. *For each $\varepsilon>0$ there exists n_ε sufficiently large that if $x,x'\in\Sigma\cup\Sigma_*$ satisfy $x_i=x'_i$ for $i=0,1,\dots,n_\varepsilon$ then*

(6.1) $$N_*(a,x) \leqslant N_*(a+\varepsilon, x')$$

for all $a \in \mathbf{R}$.

Proof. If $x_i = x_i'$ for $i=0,1,\dots,n_\varepsilon$, some $n_\varepsilon \geqslant 0$, then there is a natural, 1–1 correspondence between those y such that $\sigma^n y = x$ and those y' such that $\sigma^n y' = x'$. In particular, $y \leftrightarrow y'$ iff $y_i = y_i'$ for $i=0,1,\dots,n+n_\varepsilon$.

Recall that the function g_* in (3.1) satisfies $\mathrm{var}_m g_* = 0$, i.e., $g_*(x)$ is a function of only the first m coordinates of x. Choose $n_\varepsilon \geqslant m$; then if $y \leftrightarrow y'$ as in the preceding paragraph, $g_*(y) = g_*(y')$. Consequently, the series (3.1) defining $N_*(a,x)$ and $N_*(a+\varepsilon, x')$ only differ in terms where $S_n f_*(y) \leqslant a$ but $S_n f_*(y') > a+\varepsilon$, or vice versa. Now if $y \leftrightarrow y'$, $\sigma^n y = x$, and $\sigma^n y' = x'$, then

$$|S_n f_*(y) - S_n f_*(y')| \leqslant \sum_{m=n_\varepsilon}^{n+n_\varepsilon} \mathrm{var}_m(f_*),$$

and since $\mathrm{var}_m(f_*)$ decreases exponentially to zero, if n_ε is sufficiently large then $\Sigma_{m=n_\varepsilon}^{\infty} \mathrm{var}_m(f_*) < \varepsilon$. Therefore, all terms in the series (3.1) defining $N(a,x)$ are included in the series defining $N(a+\varepsilon, x')$, and because all terms in these series are nonnegative, (6.1) follows. □

Proof of Theorem 4. Consider the renewal equation (3.2); iteration yields

(6.2)
$$N_*(a,x) = \sum_{\substack{y:\sigma_*^n y = x \\ \sigma_*^{n-1} y \neq \xi}} N_*(a - S_n f_*(y), y) + \sum_{m=1}^{n-1} \sum_{\substack{y:\sigma^m y = x \\ \sigma^{m-1} y \neq \xi}} g(y)\, 1\{a - S_m f_*(y) \geqslant 0\} + g(x)\, 1\{a \geqslant 0\}.$$

Fix n large. For $a > n\|f_*\|_\infty$ the last two sets of terms in this expression do not change with a, so the asymptotic behavior of $N_*(a,x)$ is completely determined by

$$\sum_{\substack{y:\sigma_*^n y = x \\ \sigma_*^{n-1} y \neq \xi}} N_*(a - S_n f_*(y), y).$$

Observe that each y in the preceding expression is a sequence of length of least n, because $\sigma_*^{n-1} y \neq \xi$. For each such y there exists $y' \in \Sigma$ such that $y_i' = y_i$ for $i=0,1,\dots,n-1$. By Lemma 6.2, if $n \geqslant n_\varepsilon$ then

$$N_*(a-\varepsilon - S_n f_*(y), y') \leqslant N_*(a - S_n f_*(y), y) \leqslant N_*(a+\varepsilon - S_n f_*(y), y').$$

The asymptotic formula (3.3) now follows easily. Theorem 1 implies that

$$N_*(a\pm\varepsilon-S_nf_*(y),y')\sim C(y')\,e^{(a-S_nf_*(y))\delta}e^{\pm\varepsilon\delta}$$

as $a\to\infty$. Letting $\varepsilon\to 0$ one obtains (3.3). The continuity of $C(x)$ for $x\in\Sigma\cup\Sigma_*$ follows easily from Lemma 6.2. □

7. Perturbation theory for Perron-Frobenius operators

The proofs of Theorems 1–3 will rely heavily on analyticity properties of the map $z\to\mathscr{L}_{zf}$, $z\in\mathbf{C}$. In this section the most important such properties are summarized.

Henceforth $f\in\mathscr{F}_\varrho$ is a fixed, real-valued function such that S_mf is strictly positive for some m. The quantities $\mathscr{L}_{zf}$, λ_{zf}, h_{zf}, ν_{zf}, etc., will be abbreviated $\mathscr{L}_z$, λ_z, h_z, ν_z, etc.

The spectrum of $\mathscr{L}_z$ for $z\notin\mathbf{R}$ was partially described by Pollicott [23].

THEOREM B. *Suppose* $z\in\mathbf{C}-\mathbf{R}$.

(a) *If for some* $a\in\mathbf{R}$ *the function* $(\mathrm{Im}(z)f-a)/2\pi$ *is cohomologous to an integer-valued function, then* $e^{ia}\lambda_{\mathrm{Re}(z)}$ *is a simple eigenvalue of* $\mathscr{L}_z$, *and the rest of the spectrum is contained in a disc centered at zero of radius less than* $\lambda_{\mathrm{Re}(z)}$.

(b) *Otherwise, the entire spectrum of* $\mathscr{L}_z$ *is contained in a disc centered at zero of radius less than* $\lambda_{\mathrm{Re}(z)}$.

In case (a), $\mathrm{Im}(z)f=ia+2\pi i\varphi+i\gamma-i\gamma\circ\sigma$ for some integer-valued φ and real-valued γ. It is easily verified that the eigenvectors of $\mathscr{L}_z$ and $\mathscr{L}_z^*$ corresponding to the eigenvalue $e^{ia}\lambda_{\mathrm{Re}(z)}$ are $e^{-i\gamma}h_{\mathrm{Re}(z)}$ and $e^{i\gamma}\nu_{\mathrm{Re}(z)}$. Observe that if f is a nonlattice function then case (a) can occur only if the constant a is irrational. Consequently, if f is nonlattice then $\lambda_{\mathrm{Re}(z)}$ is in the spectrum of $\mathscr{L}_z$ iff $\mathrm{Im}(z)=0$.

In applying Theorem B the following result is useful.

LEMMA 7.1. *One of the following three cases obtains.*

(a) f *is cohomologous to a constant function.*

(b) *There do not exist* $a\in\mathbf{R}$, $b>0$ *such that* $b^{-1}f-a$ *is cohomologous to an integer-valued function.*

(c) *There exists a maximal* $b>0$ *such that for some* a, $b^{-1}f-a$ *is cohomologous to an integer-valued function. Moreover, if* $b_*, a_*\in\mathbf{R}$ *are such that* $b_*^{-1}f-a_*$ *is cohomologous to an integer-valued function, then* $b_*|b$.

Proof. Let G be the additive subgroup of $\mathbf{R}$ generated by $\{S_nf(x)-S_nf(y)\colon \sigma^nx=x,$

$\sigma^n y=y$, $n=1,2,\ldots\}$. If f is cohomologous to a constant then $G=\{0\}$. If f is cohomologous to $\varphi-c$, where c is a constant and φ takes values in a discrete subgroup H of $\mathbf{R}$, then G is a subgroup of H, hence discrete. Therefore, if neither (a) nor (b) obtains then G is a nontrivial discrete subgroup of $\mathbf{R}$. By rescaling f we may assume that $G=\mathbf{Z}$ or $G=\{0\}$.

There exists a constant $c\in[0,1)$ such that $f-c$ is cohomologous to a function valued in G. To see this, observe that for each $m=1,2,\ldots$ there exists $c_m\in[0,1)$ such that $S_{m!}f(x)-m!\,c_m\in G$ for all $x\in\Sigma$ such that $\sigma^{m!}x=x$. There exists a subsequence of c_m converging to some $\bar{c}\in[0,1]$; by construction $S_n f(x)-n\bar{c}\in G$ whenever $\sigma^n x=x$, $n=1,2,\ldots$. By Proposition 2 of [23] ((i)$\Rightarrow$(iii)), $f-\bar{c}$ is cohomologous to an integer-valued function. If $\bar{c}=1$, it may be replaced by $c=0$.

This proves that if neither (a) nor (b) holds, then (c) must. □

In the remainder of this section we shall study the operator-valued function $z\to(I-\mathscr{L}_z)^{-1}$.

The operator-valued function $z\to\mathscr{L}_z$ is an entire, holomorphic function of z; its derivative is given by

$$[(d/dz)\,\mathscr{L}_z]\,g(x)=\sum_{y:\,\sigma y=x} f(y)\,e^{zf(y)}g(y)=\mathscr{L}_z(fg)(x). \tag{7.1}$$

Therefore, if $(I-\mathscr{L}_z)^{-1}$ exists (as a bounded, linear operator on $\mathscr{F}_\varrho$) for all z in some open $D\subset\mathbf{C}$ then $z\to(I-\mathscr{L}_z)^{-1}$ is holomorphic in D. Now Proposition 2.1 implies that $\lambda_z<1$ for all $z<-\delta$, and Theorem B implies that the spectral radius of $\mathscr{L}_z$ is $\leq\lambda_{\mathrm{Re}(z)}$ for all $z\in\mathbf{C}$; consequently, if $\mathrm{Re}(z)<-\delta$ then the spectral radius of $\mathscr{L}_z$ is less than 1. This proves

PROPOSITION 7.1. *$z\to(I-\mathscr{L}_z)^{-1}$ is holomorphic in the half-plane* $\mathrm{Re}(z)<-\delta$.

Next we shall investigate the singularity of $z\to(I-\mathscr{L}_z)^{-1}$ at the point $z=-\delta$. Theorem A and standard results in regular perturbation theory ([8], Sections 7.1, 4.3) imply that the functions $z\to\lambda_z$, $z\to h_z$, $z\to\nu_z$ extend to holomorphic functions in a neighborhood N of the real line, such that

$$\lambda_z\neq 0,\quad \mathscr{L}_z h_z=\lambda_z h_z,\quad \mathscr{L}_z^*\nu_z=\lambda_z\nu_z,\quad \nu_z(h_z)=\nu_0(h_z)=1 \tag{7.2}$$

for $z\in N$. (The function $z\to\nu_z$ takes values in the dual space $\mathscr{F}_\varrho^*$, not the space of Borel measures; it is holomorphic in the weak sense that for each $g\in\mathscr{F}_\varrho$ the map $z\to\nu_z(g)$ is holomorphic.)

Define operators $\mathscr{L}_z'$ and $\mathscr{L}_z''$ for $z\in N$ by

(7.3) $$\mathscr{L}'_z g = \lambda_z \nu_z(g) h_z, \quad g \in \mathscr{F}_\varrho,$$

(7.4) $$\mathscr{L}''_z = \mathscr{L}_z - \mathscr{L}'_z.$$

By the results of the preceding paragraph, $z \to \mathscr{L}'_z$ and $z \to \mathscr{L}''_z$ are holomorphic, operator-valued functions in N. For each $z \in N$, $\mathscr{L}'_z$ maps $\mathscr{F}_\varrho$ onto the one-dimensional subspace generated by h_z, whereas $\mathscr{L}''_z$ maps $\mathscr{F}_\varrho$ onto the complementary subspace $\{g \in \mathscr{F}_\varrho : \nu_z(g) = 0\}$. Therefore,

(7.5) $$\mathscr{L}^n_z = (\mathscr{L}'_z)^n + (\mathscr{L}''_z)^n$$

for $n=1,2,\dots$. Notice that $\lambda_z^{-1}\mathscr{L}'_z$ is idempotent, so $(\mathscr{L}'_z)^n g = \lambda_z^n \nu_z(g) h_z$.

Theorem A (Section 1) implies that for $z \in \mathbf{R}$ the spectral radius of $\mathscr{L}''_z$ is less than λ_z. Since $z \to \mathscr{L}''_z$ is analytic, the spectral radius of $\mathscr{L}''_z$ is a lower semicontinuous function of z. Consequently, there is a neighborhood U of $z=-\delta$ such that the spectral radius of $\mathscr{L}''_z$ is less than $1-2\varepsilon$ for all $z \in U \cap N$, for some $\varepsilon > 0$. It now follows from the spectral radius formula that for some $n_* \geqslant 1$, $\|(\mathscr{L}''_z)^n\| \leqslant (1-\varepsilon)^n$ for all $n \geqslant n_*$ and $z \in U \cap N$. Hence,

(7.6) $$z \to \sum_{n=0}^{\infty} (\mathscr{L}''_z)^n = (I - \mathscr{L}''_z)^{-1}$$

is a holomorphic, operator-valued function of $z \in U \cap N$. Together with (7.3)–(7.5), this proves

PROPOSITION 7.2. *The function $z \to (I-\mathscr{L}_z)^{-1}$ has a simple pole at $z=-\delta$. In particular, for each $g \in \mathscr{F}_\varrho$*

(7.7) $$(I-\mathscr{L}_z)^{-1} g = (1-\lambda_z)^{-1} \nu_z(g) h_z + (I-\mathscr{L}''_z)^{-1} g$$

for z in some punctured neighborhood of $z=-\delta$.

The regularity properties of $z \to (I-\mathscr{L}_z)^{-1}$ on the rest of the line $\mathrm{Re}(z) = -\delta$ depend on which case of Lemma 7.1 obtains. The case where f is cohomologous to a constant we shall ignore. Suppose there do not exist constants $a \in \mathbf{R}$, $b > 0$, such that $b^{-1}f - a$ is cohomologous to an integer-valued function. Then by Theorem B the spectral radius of $\mathscr{L}_z$ is less than $\lambda_{\mathrm{Re}(z)}$ if $\mathrm{Im}(z) \neq 0$. Since the spectral radius of $\mathscr{L}_z$ is lower semicontinuous in z, it follows that for each z_* on the line $\mathrm{Re}(z) = -\delta$ except $z_* = -\delta$ there is a neighborhood of z_* in which $z \to (I-\mathscr{L}_z)^{-1}$ is holomorphic.

Suppose next that f is nonlattice but that $b^{-1}f - a$ is cohomologous to an integer-valued function. Assume that $b>0$ is maximal in the sense of Lemma 7.1(c). Then ab is irrational, so by Theorem B, $\lambda_{\mathrm{Re}(z)}$ is in the spectrum of $\mathscr{L}_z$ iff $\mathrm{Im}(z)=0$. Consequently, $z\to(I-\mathscr{L}_z)^{-1}$ is holomorphic in a neighborhood of every z on the line $\mathrm{Re}(z)=-\delta$ except $z=-\delta$. This proves

PROPOSITION 7.3. *If f is nonlattice then the function $z\to(I-\mathscr{L}_z)^{-1}$ is holomorphic in a neighborhood of every z on the line* $\mathrm{Re}(z)=-\delta$ *except* $z=-\delta$.

Next, suppose that f is integer-valued and that f is not cohomologous to any function valued in a proper subgroup of the integers. Then $\mathscr{L}_z$ is $2\pi i$-periodic in z, so the pole at $z=-\delta$ is repeated at $z=-\delta+2\pi im$, $m\in\mathbf{Z}$. Theorem B implies that $1=\lambda_{-\delta}$ is not in the spectrum of $\mathscr{L}_z$ when $\mathrm{Re}(z)=-\delta$ and $\mathrm{Im}(z)/2\pi$ is not an integer. Hence,

PROPOSITION 7.4. *If f is integer-valued but not cohomologous to any function valued in a proper subgroup of the integers then $z\to(I-\mathscr{L}_z)^{-1}$ is $2\pi i$-periodic, and holomorphic at every z on* $\mathrm{Re}(z)=-\delta$ *such that* $\mathrm{Im}(z)/2\pi$ *is not an integer.*

8. Fourier analysis of the renewal equation

In this section we shall analyze the asymptotic behavior of $N(a,x)$ as $a\to\infty$ by means of its Fourier transform. We begin with the lattice case, as it is easier.

Proof of Theorem 2. Suppose that f is an integer-valued function but that f is not cohomologous to a function taking values in a proper subgroup of the integers. In this case $N(a,x)$ is a step function in a with discontinuities only at integer values of a. Recall that $N(a,x)=0$ for $a\leqslant 0$ and that $N(a,x)=O(e^{Ca})$ for some $C<\infty$ as $a\to\infty$ (see Lemma 8.1); consequently, the Fourier-Laplace transform

$$\hat{N}(z,x)=\sum_{n=-\infty}^{\infty} e^{nz}N(n,x) \tag{8.1}$$

is well-defined and analytic for z in a half-plane $\mathrm{Re}\,z<-C$. The renewal equation (2.2) transforms as

$$\hat{N}(z,x)=\mathscr{L}_z\hat{N}(z,x)+g(x)/(1-e^z)$$

for $\mathrm{Re}(z)<0$. By Proposition 7.1 $(I-\mathscr{L}_z)^{-1}$ is defined and holomorphic in the half-plane $\mathrm{Re}(z)<-\delta$, hence in the half-plane $\mathrm{Re}(z)<\min(-\delta,-C)$ the functional equation

$$\hat{N}(z,x)=(1-e^z)^{-1}(I-\mathscr{L}_z)^{-1}g(x)$$

holds. Since the right hand side of this equation is analytic in $\mathrm{Re}(z)<-\delta$, $\hat{N}(z,x)$ admits an extension to $\mathrm{Re}(z)<-\delta$. Since the coefficients $N(n,x)$ in the series (8.1) are nonnegative, it follows that the series converges uniformly and absolutely for $\mathrm{Re}(z)<-\delta$.

According to Propositions 7.2 and 7.4, $(I-\mathscr{L}_z)^{-1}g(x)$ has an isolated singularity at $z=-\delta$ but is regular (holomorphic) at each $z=-\delta+i\theta$, $0<|\theta|\leq\pi$. The singularity at $z=-\delta$ is $(1-\lambda_z)^{-1}\nu_z(g)\,h_z(x)$; since $\nu_z(g)$ and $h_z(x)$ are continuous at $z=-\delta$, the singularity is a simple pole with residue

$$\frac{\nu_{-\delta}(g)\,h_{-\delta}(x)}{(-(d/dz)\,\lambda_z)_{z=-\delta}}=\left(\frac{\int g\,d\nu_{-\delta}}{\int f\,d\mu_{-\delta}}\right)h_{-\delta}(x).$$

It follows that $\hat{N}(z,x)$ is meromorphic in $\{z\colon 0\leq\mathrm{Im}(z)\leq\pi,\ \mathrm{Re}(z)<-\delta+\varepsilon\}$ for some $\varepsilon>0$, and that the only singularity in this region is a simple pole at $z=-\delta$ with residue $C(x)$, where $C(x)$ is given by (2.7). Consequently,

$$F(z,x)\stackrel{\Delta}{=}\sum_{n=0}^{\infty}z^n\{e^{-n\delta}N(n,x)-C(x)\}$$

is holomorphic in $\{|z|<1+2\varepsilon\}$ for some $\varepsilon>0$; F is also jointly continuous in z and x. Cauchy's integral formula now implies that

$$e^{-n\delta}N(n,x)-C(x)=(2\pi i)^{-1}\int_{|z|=1+\varepsilon}F(z,x)\,z^{-n-1}\,dz=O((1+\varepsilon)^{-n})$$

as $n\to\infty$, uniformly for $x\in\Sigma$. This proves (2.6). □

Proof of Theorem 3. Assume that

$$f=\varphi+\gamma-\gamma\circ\sigma$$

where φ is an integer-valued function, and that f (hence also φ) is not cohomologous to any function valued in a proper subgroup of the integers. Note that since $S_m f>0$ for some m, there exists m' such that $S_{m'}\varphi>0$. Recall that $\lambda_z=\lambda_{zf}=\lambda_{z\varphi}$ for all $z\in\mathbf{R}$ hence also for those $z\in\mathbf{C}$ where λ_z is defined by analytic continuation.

Define

$$N^*(a,x)=N(a-\gamma(x),x)$$

for $a\in\mathbf{R}$, $x\in\Sigma$. The renewal equation (2.2) for $N(a,x)$ may be rewritten as

$$N^*(a,x)=\sum_{\sigma y=x}N^*(a-\varphi(y),y)+g(x)\,1\{a\geq\gamma(x)\}.$$

Fix $\beta\in[0,1)$, and define the Fourier-Laplace transform

$$\hat{N}^*_\beta(z,x)=\sum_{n=-\infty}^{\infty}e^{nz}N^*(n+\beta,x).$$

The renewal equation implies that

$$\hat{N}^*_\beta(z,x)=(1-e^z)^{-1}(I-\mathcal{L}_{z\varphi})^{-1}(ge^{z[\gamma+1-\beta]})(x)$$

for $\mathrm{Re}(z)<-\delta$, as in the proof of Theorem 2. We may remove the singularity at $z=-\delta$ as before to obtain

$$N^*(n+\beta,x)\sim C(\beta,x)\,e^{n\delta} \tag{8.2}$$

as $n\to\infty$, uniformly for $x\in\Sigma$. This also holds uniformly for $\beta\in[0,1)$, because the functions $g(x)\exp\{z[\gamma(x)+1-\beta]\}$, $\beta=[0,1)$, are bounded in $\mathcal{F}_\varrho$. Now (8.2) clearly implies (2.8). □

The nonlattice case is more complicated because it requires the analysis of a Fourier integral rather than a Fourier series. The Fourier (integral) transform of $N(a,x)$ once again involves $(I-\mathcal{L}_z)^{-1}$. Now Proposition 7.3 states that $(I-\mathcal{L}_z)^{-1}$ is holomorphic at every z such that $\mathrm{Re}(z)=-\delta$, $\mathrm{Im}(z)\neq 0$; however, the operator norm of $(I-\mathcal{L}_z)^{-1}$ may be unbounded as $|\mathrm{Im}(z)|\to\infty$ on the line $\mathrm{Re}(z)=-\delta$. Thus, the Fourier transform of $e^{-a\delta}N(a,x)$ may behave wildly at ∞. Circumventing this problem requires an unsmoothing argument, and this in turn requires an *a priori* bound on the growth of the renewal function.

LEMMA 8.1. *There exists constants $C, a_*<\infty$, depending only on f, such that for every $g\in\mathcal{F}_\varrho$ the function $N(a,x)$ defined by* (2.1) *satisfies*

$$N(a,x)=0,\quad \forall a\leq a_*, \tag{8.3}$$

$$N(a,x)\leq C\|g\|_\infty\, e^{a\delta}. \tag{8.4}$$

Proof. Recall that there is an $n\geq 1$ such that $S_nf\geq\varepsilon>0$ on Σ. Since $f, S_2f, \ldots, S_{n-1}f$ are continuous and Σ is compact, there exists $a_*\in(-\infty,0)$ such that

$$\min_{0\leq i\leq n-1} S_if(x)>a_*,\quad \forall x\in\Sigma. \tag{8.5}$$

If $a \leqslant a_*$ then (8.5) implies that there are no nonzero terms in the series (2.1) defining $N(a,x)$; thus (8.3).

Define $G(a,x)=e^{-a\delta}N(a,x)/h_{-\delta f}(x)$. Since $h_{-\delta f}$ is bounded away from 0 and ∞, it suffices to prove that $G(a,x)\leqslant C\|g\|_\infty$ for a suitable $C<\infty$.

The renewal equation (2.2) may be rewritten as

$$G(a,x)=\sum_{y:\,\sigma^n y=x} G(a-S_nf(y),y)\,e^{-\delta S_nf(y)}(h_{-\delta f}(y)/h_{-\delta f}(x)) + e^{-a\delta}\sum_{i=0}^{n-1}\sum_{\sigma^i y=x} g(y)\,(1\{a\geqslant S_if(y)\}/h_{-\delta f}(x)). \tag{8.6}$$

Observe that

$$\sum_{y:\,\sigma^n y=x} e^{-\delta S_nf(y)}(h_{-\delta f}(y)/h_{-\delta f}(x))=(\mathscr{L}^n_{-\delta f}h_{-\delta f})(x)/h_{-\delta f}(x)=1, \tag{8.7}$$

since $h_{-\delta f}$ is the eigenfunction of $\mathscr{L}_{-\delta f}$ corresponding to the eigenvalue $\lambda_{-\delta f}=1$. If n is sufficiently large that $S_nf(y)\geqslant\varepsilon>0$ for all $y\in\Sigma$, and if

$$\bar{G}(a)=\sup_{a'\leqslant a,\,x\in\Sigma} G(a',x),$$

then it follows from (8.6) and (8.7) that for all $a\in\mathbf{R}$,

$$\bar{G}(a)\leqslant\bar{G}(a-\varepsilon)+C'\|g\|_\infty e^{-a\delta} \tag{8.8}$$

where

$$C'=\sum_{i=0}^{n-1} l^i\Big/\min_{y\in\Sigma} h_{-\delta f}(y)<\infty.$$

Now (8.8) implies that $\bar{G}(a)\leqslant C\|g\|_\infty$ for a suitable constant $C<\infty$, proving (8.4). □

To prove that $e^{-a\delta}N(a,x)\to C(x)$ we will show that it suffices to prove that a suitably smoothed version of $e^{-a\delta}N(a,x)$ converges to $C(x)$, then accomplish the latter by Fourier analysis. For the smoothing we will use a continuous probability density $k(t)=k(-t)$ whose Fourier transform $\hat{k}(i\theta)=\int e^{i\theta t}k(t)\,dt=\hat{k}(-i\theta)$ is nonnegative, C^∞, and has compact support. Let P be the set of all such probability densities. (Note: $P\neq\emptyset$. To see this let $\hat{k}=\hat{k}_1 * \hat{k}_1$, where $\hat{k}_1$ is an even, C^∞, compactly supported function, suitably normalized. Then $k\in P$.)

LEMMA 8.2. *To prove Theorem* 1 *it suffices to prove that for each* $k\in P$

$$\lim_{a\to\infty}\int_{-\infty}^{\infty} k(a-t)\, e^{-\delta t} N(t,x)\, dt = C(x) \tag{8.9}$$

uniformly for $x\in\Sigma$.

Proof. Fix $\varepsilon>0$ (small). Choose $k\in P$ such that $\int_{-\varepsilon}^{\varepsilon} k(t)\,dt>1-\varepsilon$ (such a k exists because if $k(t)\in P$ then $ck(ct)\in P$ for all $c>0$). By Lemma 8.1, there exists $C<\infty$ such that $e^{-\delta t}N(t,x)\leqslant C$ for all $t\in\mathbf{R}$; consequently, if (8.9) holds then for sufficiently large a and all $x\in\Sigma$

$$\left|\int_{a-\varepsilon}^{a+\varepsilon} k(a-t)\, e^{-\delta t} N(t,x)\, dt - C(x)\right| < C'\varepsilon,$$

for a suitable $C'<\infty$ independent of x. Since $N(t,x)$ is nondecreasing in t,

$$\int_{a}^{a+2\varepsilon} k(a+\varepsilon-t)\, e^{-\delta t} N(t,x)\, dt \geqslant (1-\varepsilon)\, e^{-2\delta\varepsilon} e^{-\delta a} N(a,x)$$

and

$$\int_{a-2\varepsilon}^{a} k(a-\varepsilon-t)\, e^{-\delta t} N(t,x)\, dt \leqslant e^{2\delta\varepsilon} e^{-\delta a} N(a,x).$$

Letting $\varepsilon\to 0$ shows that $e^{-\delta a}N(a,x)\to C(x)$ uniformly for $x\in\Sigma$ as $a\to\infty$. □

Proof of Theorem 1. By Lemma 8.2 it suffices to prove (8.9) for arbitrary $k\in P$. Recall that $N(a,\cdot)\equiv 0$ for $a\leqslant 0$ and that $e^{-\delta a}N(a,x)$ is uniformly bounded for $a\in\mathbf{R}$, $x\in\Sigma$. Hence, to prove (8.9) it suffices to show that

$$\lim_{a\to\infty}\int_{-\infty}^{\infty} \{k(a-t)+k(-a-t)\}\, e^{-\delta t} N(t,x)\, dt = C(x) \tag{8.10}$$

uniformly for $x\in\Sigma$.

Define the Fourier-Laplace transform

$$F(z,x) = \int_{-\infty}^{\infty} e^{za-\delta a} N(a,x)\, da. \tag{8.11}$$

As in the proof of Theorem 1, the renewal equation (2.2) and the analyticity of $z \to (I-\mathscr{L}_z)^{-1}$ in the half-plane $\mathrm{Re}(z) < -\delta$ imply that

$$F(z,x) = -(I-\mathscr{L}_{z-\delta})^{-1} g(x)/(z-\delta)$$

for $\mathrm{Re}(z)<0$ (observe that Lemma 8.1 implies that the integral (8.11) converges absolutely and uniformly in each half-plane $\mathrm{Re}(z)<-\varepsilon<0$). By Proposition 7.3, $F(z,x)$ is holomorphic at every z on the line $\mathrm{Re}(z)=0$ except $z=0$. By Proposition 7.2, $F(z,x)$ has a simple pole at $z=0$ with residue $-C(x)$, where $C(x)$ is given by (2.5). Therefore,

$$F(z,x) = -C(x)/z + G(z,x),$$

where $G(z,x)$ is holomorphic in a region containing the closed half-plane $\mathrm{Re}(z) \geqslant 0$.

The monotone convergence theorem implies that

$$\int_{-\infty}^{\infty} \{k(a-t)+k(-a-t)\}\, e^{-\delta t} N(t,x)\, dt = \lim_{s \downarrow 0} \int_{-\infty}^{\infty} \{k(a-t)+k(-a-t)\}\, e^{-(\delta+s)t} N(t,x)\, dt.$$

For each $s>0$, $e^{-(\delta+s)t}N(t,x)$ is a nonnegative, integrable function (relative to dt), by Lemma 8.1. Furthermore, $k(t)$ is the (inverse) Fourier transform of a nonnegative, integrable function $\hat{k}(i\theta)$. Consequently, the Parseval relation ([3], Chapter XV, equation (3.2)) implies that

$$(8.12) \qquad \int_{-\infty}^{\infty} \{k(a-t)+k(-a-t)\}\, e^{-(\delta+s)t} N(t,x)\, dt = \int_{-\infty}^{\infty} 2\cos(\theta a)\, \hat{k}(i\theta)\, F(-s+i\theta,x)\, d\theta/2\pi.$$

Since the left hand side is real, we may ignore the imaginary part of the integrand on the right hand side. Using the representation (8.12) together with the fact that $\hat{k}(i\theta)$ is real we obtain

$$\int_{-\infty}^{\infty} \{k(a-t)+k(-a-t)\}\, e^{-(\delta+s)t} N(t,x)\, dt = \int_{-\infty}^{\infty} 2\cos(a\theta)\, \hat{k}(i\theta)\, C(x)\, s(s^2+\theta^2)^{-1}\, d\theta/2\pi$$
$$+ \int_{-\infty}^{\infty} 2\cos(a\theta)\hat{k}(i\theta)\,\mathrm{Re}\, G(-s+i\theta,x)\, d\theta/2\pi \to C(x) + \int_{-\infty}^{\infty} \cos(a\theta)\hat{k}(i\theta)\,\mathrm{Re}\, G(i\theta,x)\, d\theta/\pi$$

as $s \downarrow 0$. (The measures $\pi^{-1}s(s^2+\theta^2)\, d\theta$ converge weakly to the delta function at 0 as $s \downarrow 0$, and $\hat{k}(0)=1$ because k is a probability density.)

It remains to show that the last integral converges to zero as $a \to \infty$, uniformly for $x \in \Sigma$. Recall that $\hat{k}(i\theta)$ has compact support, say $[-A,A]$, and is C^{∞}, also that $G(i\theta,x)$ is analytic in a neighborhood of $[-A,A]$ and continuous in x. The Cauchy integral formula for derivatives implies that $(d/d\theta)\,G(i\theta,x)$ is uniformly continuous, hence bounded on $[-A,A]\times\Sigma$. Integrating by parts gives

$$\int_{-A}^{A}\cos(a\theta)\,\hat{k}(i\theta)\operatorname{Re}G(i\theta,x)\,d\theta=-a^{-1}\int_{-A}^{A}\sin(a\theta)\,(d/d\theta)\,\{\hat{k}(i\theta)\operatorname{Re}G(i\theta,x)\}\,d\theta;$$

clearly, this converges to zero as $a\to\infty$, uniformly for $x\in\Sigma$. This proves (8.10). □

Part II. Applications to discrete groups

9. Symbolic dynamics for Schottky groups

Let $Q_1, Q_{-1}, Q_2, Q_{-2}, \ldots, Q_k, Q_{-k}$ be $2k$ ($k\geqslant 2$) mutually exterior (nonintersecting) circles in the plane, and for each $j=1,2,\ldots,k$ let T_j be a linear fractional transformation mapping the exterior of Q_{-j} onto the interior of Q_j. The group Γ generated by $T_1, T_2, \ldots, T_k$ is called a *Schottky group* ([4], [15]). The region $\mathcal{R}$ exterior to all $2k$ circles is a fundamental region for the group.

The Schottky group Γ enjoys a very simple and transparent symbolic dynamics, because it is a free group on the k generators $T_1, T_2, \ldots, T_k$. Consider the set Σ_* of finite sequences from the alphabet $\{T_1, T_1^{-1}, T_2, T_2^{-1}, \ldots, T_k^{-1}\}$ with all transitions allowed except $T_iT_i^{-1}$ and $T_i^{-1}T_i$, $i=1,\ldots,k$. There is a natural 1–1 correspondence between Γ and Σ_* in which the group identity corresponds to the empty sequence ξ and the shift σ_* corresponds to a multiplication by one of the symbols $T_i^{\pm 1}$. Henceforth we will not always distinguish between Γ and Σ_*; $x_1x_2\ldots x_n$ may denote a sequence in Σ_* or the corresponding element of Γ.

To complete the symbolic dynamics we will extend the correspondence just described to a map from the set Σ of *infinite* sequences (with allowable transitions) to the limit set Λ of Γ. To accomplish this we will show that for any $x=x_1x_2\ldots\in\Sigma$ and any $z\in\mathcal{R}$, $\lim_{n\to\infty}(x_1x_2\ldots x_n)(z)$ exists and is independent of $z\in\mathcal{R}$.

For each $\gamma=T_i^{\pm 1}$, $i=1,2,\ldots,k$, let D_γ be the open disc interior to $Q_\gamma=Q_{\pm i}$; for $x_1x_2\ldots x_n\in\Sigma_*-\{\xi\}$, let

$$D_{x_1x_2\ldots x_n}=x_1D_{x_2x_3\ldots x_n}=\ldots=x_1\ldots x_{n-1}D_{x_n}. \tag{9.1}$$

Notice that

$$\bar{D}_{x_1x_2\ldots x_{n+m}}\subset D_{x_1x_2\ldots x_n}\quad\text{if}\quad m\geqslant 1; \tag{9.2}$$

$$\bar{D}_{x_1x_2\ldots x_n}\cap\bar{D}_{x_1'x_2'\ldots x_n'}=\varnothing\quad\text{unless}\quad x_i=x_i',\ i=1,\ldots,n; \tag{9.3}$$

(9.4) $$x_1x_2\dots x_n\mathcal{R}\subset D_{x_1x_2\dots x_n}$$

(compare with [7], Section 5.2).

LEMMA 9.1. *There exist constants $C<\infty$, $0<\varrho<1$ such that for each $x_1x_2\dots x_n\in\Sigma_*$*

(9.5) $$\operatorname{diam}(D_{x_1x_2\dots x_n})\leq C\varrho^n.$$

The proof will be given at the end of the section.

It is now apparent from (9.2)–(9.5) that for each $x=x_1x_2\ldots\in\Sigma$ the regions $x_1x_2\dots x_n\mathcal{R}$ shrink to a single point $x\in\Lambda$ as $x\to\infty$, and that the induced map $\Sigma\to\Lambda$ is a homeomorphism. Furthermore, (9.5) implies that any Lipschitz continuous function on Λ pulls back to a Hölder continuous ($\mathcal{F}_\varrho$) function on Σ.

Consider now the shift $\sigma:\Sigma\to\Sigma$; let $F:\Lambda\to\Lambda$ be the corresponding map on Λ. Following Bowen [2] we will call $F=F_\Gamma$ the *Nielsen map* for Γ (cf. [18]). Notice that for $z\in\Lambda\cap D_{\pm i}$,

$$F(z)=T_i^{\mp1}z,$$

so F is continuously differentiable and $F'(z)\neq0$ at every $z\in\Lambda$. Define

(9.6) $$\bar{f}(z)=\log|F'(z)|,\quad z\in\Lambda,$$

and let $f:\Sigma\to\mathbf{R}$ be the Hölder continuous pullback of $\bar{f}$. We will call f (or $\bar{f}$) the *distortion function* of Γ.

LEMMA 9.2. *There exists $n\geq1$ such that S_nf is strictly positive on Σ.*

The proof will be given at the end of the section.

Any measure on the sequence space Σ induces a corresponding measure on the limit set Λ (and vice-versa). Bowen [2] proved that the Gibbs measure $\mu_{-\delta f}$ on Σ (here $\delta>0$ is as in Proposition 2.1) induces a measure on Λ that is equivalent to δ-dimensional Hausdorff measure on Λ, and that the δ-dimensional Hausdorff measure of Λ is finite and strictly positive. Series [27] subsequently proved that the measure induced by $\nu_{-\delta f}$ is actually a scalar multiple of the δ-dimensional Hausdorff measure on Λ.

Proof of Lemmas 9.1–9.2. The group Γ is discontinuous at ∞, so every element $\gamma\in\Gamma$ has an isometric circle C_γ, and for each $\varepsilon>0$ only finitely many C_γ have radii larger than ε (cf. [15], IV.1 D). Moreover, for each $\gamma\neq$identity, $\gamma(\infty)$ lies in the interior of $C_{\gamma^{-1}}$.

Now (9.4) implies that $\gamma(\infty)\in D_\gamma$, so it follows from (9.2) that all but finitely many of the isometric circles C_γ lie in $\bigcup_{j=1}^{k}(\text{interior}(Q_j)\cup\text{interior}(Q_{-j}))$.

Choose n sufficiently large that each isometric circle $C_{x_1x_2\ldots x_n}$, $x_1x_2\ldots x_n\in\Sigma_*$, lies in interior $(Q_{\pm j})$ for some j. Then for every sequence $x_1\ldots x_n$ of length n and every $\pm i$, $x_1x_2\ldots x_n Q_{\pm i}\subset\text{interior}(C_{(x_1\ldots x_n)^{-1}})$; but since $x_1x_2\ldots x_n\mathcal{R}\subset D_{x_1}$ and each $Q_{\pm i}\subset\partial\mathcal{R}$, it follows that $x_1\ldots x_n Q_{\pm i}\subset\bar{D}_{x_1}$. Hence $C_{(x_1\ldots x_n)^{-1}}\subset D_{x_1}$. Thus, whenever x_nx_{n+1} is an allowable transition (i.e., $x_nx_{n+1}\neq\text{id}$), $\bar{D}_{x_{n+1}}$ lies in the exterior of $C_{x_1x_2\ldots x_n}$.

The derivative of a linear fractional transformation in absolute values is <1 outside the isometric circle. Consequently, there is an $\alpha<1$ such that for each sequence $x_1x_2\ldots x_n\in\Sigma_*$ of length n the derivative of $x_1\ldots x_n$ in absolute value is $\leqslant\alpha$ on $\bigcup_{\gamma\neq x_n^{-1}}\bar{D}_\gamma$. It follows by an easy argument that

$$\max_{x_1x_2\ldots x_{kn+1}\in\Sigma_*}\operatorname{diam} D_{x_1\ldots x_{kn+1}}\leqslant C\alpha^k,\quad k=1,2,\ldots,\tag{9.7}$$

for a suitable $C<\infty$. The result (9.5) follows from (9.7).

Finally, consider the Nielsen map F. On $\Lambda\cap D_{x_1x_2\ldots x_n}$, $F\circ F\circ\ldots\circ F$ (n times) coincides with $(x_1x_2\ldots x_n)^{-1}$ (cf. (9.6)), and maps $\Lambda\cap D_{x_1x_2\ldots x_n}$ into $\bigcup_{\gamma\neq x_n^{-1}}\bar{D}_\gamma$. Consequently, the derivative of $F\circ F\circ\ldots\circ F$ in absolute value is $\geqslant 1/\alpha>1$ on Λ. Lemma 9.2 now follows from the definition of f. □

10. Symbolic dynamics for Fuchsian groups

Series [27] has developed "symbolic dynamics" for a large class of nonelementary, finitely generated Fuchsian groups. In this section we give a resumé of her results.

A *Fuchsian group* is a discrete group Γ of linear fractional transformations, each mapping the unit disc D onto itself; it is *nonelementary* if its limit set Λ is infinite. A nonelementary Fuchsian group is necessarily nonabelian. A group Γ is *finitely generated* with generating set Γ_0 if every element of Γ is a finite product of elements of Γ_0. Each finite product of elements of Γ_0 yielding the identity in Γ is called a *relation*; the set of relations and their inverses is itself a free group.

A *nonexceptional* Fuchsian group is a nonelementary, finitely generated Fuchsian group satisfying at least one of the following conditions:

$$D/\Gamma \text{ is not compact};\tag{10.1}$$

$$|\Gamma_0|\geqslant 5 \text{ and every nontrivial relation has length }\geqslant 5;\tag{10.2}$$

(10.3) at least 3 of the generating relations have length ≥ 7.

Series proved that if Γ is a nonexceptional Fuchsian group then every element of Γ has a canonical representation as a shortest word in the generators Γ_0, and that the rules governing these representations are of finite type. Consequently, there is a 1–1 correspondence between Γ and Σ_*, where Σ_* is the set of finite sequences with admissible transitions from some finite alphabet.

Assume henceforth that Γ is a nonexceptional Fuchsian group with no parabolic elements. (A linear fractional transformation is called *parabolic* if it has only one fixed point.) Series proved that the canonical bijection $\Gamma \leftrightarrow \Sigma_*$ extends to a continuous map $\Sigma \to \Lambda$, which is onto and one-to-one except at a countable number of points where it is two to one. (Here Λ is the limit set of Γ and Σ the set of infinite sequences with the same transition rules as Σ_*.) If $z \in \mathrm{Int}(D)$ is not a fixed point, then there exist $C<\infty$, $\varrho<1$ such that whenever $x, x' \in \Sigma_*$, $x_i = x_i'$, $\forall i \leq n$ then

$$d_E(xz, x'z) \leq C\varrho^n \tag{10.4}$$

(cf. [27], Proposition 4.2; d_E is the Euclidean distance). Consequently, if $x \in \Sigma$ then $\lim_{n\to\infty}(x_1 x_2 \ldots x_n)(z) \in \Lambda$ gives a continuous surjective map $\Sigma \to \Lambda$. The inequality (10.4) lifts to Σ, and therefore Lipschitz continuous functions on Λ lift to Hölder continuous functions on Σ.

The shift $\sigma: \Sigma \to \Sigma$ induces a map $F: \Lambda \to \Lambda$, which we will again call the *Nielsen map*. This map is C^1 except (perhaps) at a finite set of points. Define

$$\tilde{f}(z) = \log|F'(z)|, \quad z \in \Lambda, \tag{10.5}$$

and let f be the Hölder continuous pullback to Σ. We will call f (or $\tilde{f}$) the *distortion function* for Γ. Series proved that there is an $n \geq 1$ such that $S_n f$ is strictly positive on Σ.

Finally, let $\delta > 0$ be the unique real number such that $\lambda_{-\delta f} = 1$ (Proposition 2.1) and $\nu_{-\delta f}$ the corresponding eigenmeasure. Series proved that δ is the Hausdorff dimension of Λ, and that $\nu_{-\delta f}$ is a scalar multiple of the δ-dimensional Hausdorff measure on Λ (cf. also [22], [31]). She calls $\nu_{-\delta f}$ the "Patterson measure".

11. The geodesic flow

Let Γ be a Schottky group. Γ may be regarded as a group of isometries of the hyperbolic space H^{d+1}, where $d=1$ if Γ is Fuchsian and $d=2$ otherwise (H^{d+1} is the unit ball in $\mathbf{R}^{d+1}$ with the Poincaré metric). Thus, H^{d+1}/Γ is a Riemannian manifold with the induced Poincaré metric. In this section we shall discuss the geodesic flow on the unit tangent bundle $T(H^{d+1}/\Gamma)$, using the symbolic dynamics developed in Section 9. The restriction to Schottky groups is primarily for simplicity; the methods and results of this section can almost certainly be generalized to the nonexceptional Fuchsian groups of Section 10 (cf. [28], [29]). Our methods can be traced back to Artin, Hedlund, and Morse (cf. [28]). However, our main result, Theorem 8, seems not to have been noticed before.

For definiteness, let Γ be a Fuchsian Schottky group generated by $T_1^{\pm 1}, T_2^{\pm 1}, \ldots, T_k^{\pm 1}$ as in Section 9. The region $\mathcal{R}$ inside the unit disc D but exterior to each of the circles $Q_1, Q_{-1}, \ldots, Q_{-k}$ is a fundamental polygon for Γ. The images $\gamma\mathcal{R}$, $\gamma\in\Gamma$ are nonoverlapping, and by Lemma 9.1 the *euclidean* diameters of $\gamma_n\mathcal{R}$ converge to zero as $n\to\infty$ for any sequence γ_n of distinct elements of Γ. Consequently, if $z_n, z_n'\in\gamma_n\mathcal{R}$ for each $n=1,2,\ldots$, then $\lim z_n$ exists iff $\lim z_n'$, exists, in which case $\lim z_n = \lim z_n' \in\Lambda$. On the other hand, if $x\in\Lambda$ has the expansion $x_1x_2\ldots\in\Sigma$, then for any γ the sequence $(x_1x_2\ldots x_n)\gamma\mathcal{R}$ of polygons converges to the point $x\in\Lambda$ as $n\to\infty$.

Consider now the set of geodesics in H^2; keep in mind that the geodesics on the Riemann surface H^2/Γ are in 1–1 correspondence with the geodesics in H^2 that enter $\mathcal{R}$ (plus k additional geodesics corresponding to the $2k$ circles bounding $\mathcal{R}$, which we will ignore.) Each geodesic in H^2 is determined (up to a translation in time) by its endpoints on the unit circle S^1. Consider a geodesic (x, y) whose left and right endpoints x and y are both in the limit set Λ; let $x\leftrightarrow x_1x_2\ldots\in\Sigma$ and $y\leftrightarrow y_1y_2\ldots\in\Sigma$. In order that the geodesic (x, y) enter $\mathcal{R}$ it is necessary and sufficient that $x_1\neq y_1$. Furthermore, if this is the case then the geodesic passes through infinitely many images $\gamma\mathcal{R}$ of $\mathcal{R}$, in the following order:

$$\ldots\to x_1x_2\mathcal{R}\to x_1\mathcal{R}\to\mathcal{R}\to y_1\mathcal{R}\to y_1y_2\mathcal{R}\to\ldots. \tag{11.1}$$

Conversely, if a geodesic in H^2 cuts through infinitely many $\gamma\mathcal{R}$ in the order specified by (11.1) then the left and right endpoints of the geodesic are necessarily x and y, by the arguments of the preceding paragraph. See Figure 2.

Now let $x, y\in\Lambda$, $x = x_1x_2\ldots$, $y=y_1y_2\ldots$. The condition that $x_1\neq y_1$ is equivalent to the condition that $x_1^{-1}y_1\neq\mathrm{id}$, i.e., that all transitions in the double-ended sequence $\ldots x_2^{-1}x_1^{-1}y_1y_2\ldots$ are allowable. Let $\bar{\Sigma}$ denote the set of all double-ended sequences with allowable transitions, and let σ denote the forward shift on $\bar{\Sigma}$. Define $g:\bar{\Sigma}\to(0,\infty)$ by setting $g(\ldots x_2^{-1}x_1^{-1}y_1y_2\ldots)$ equal to the noneuclidean length of that segment of the geodesic (x, y) lying in $\mathcal{R}$. It follows routinely from Lemma 9.1 that g is Hölder

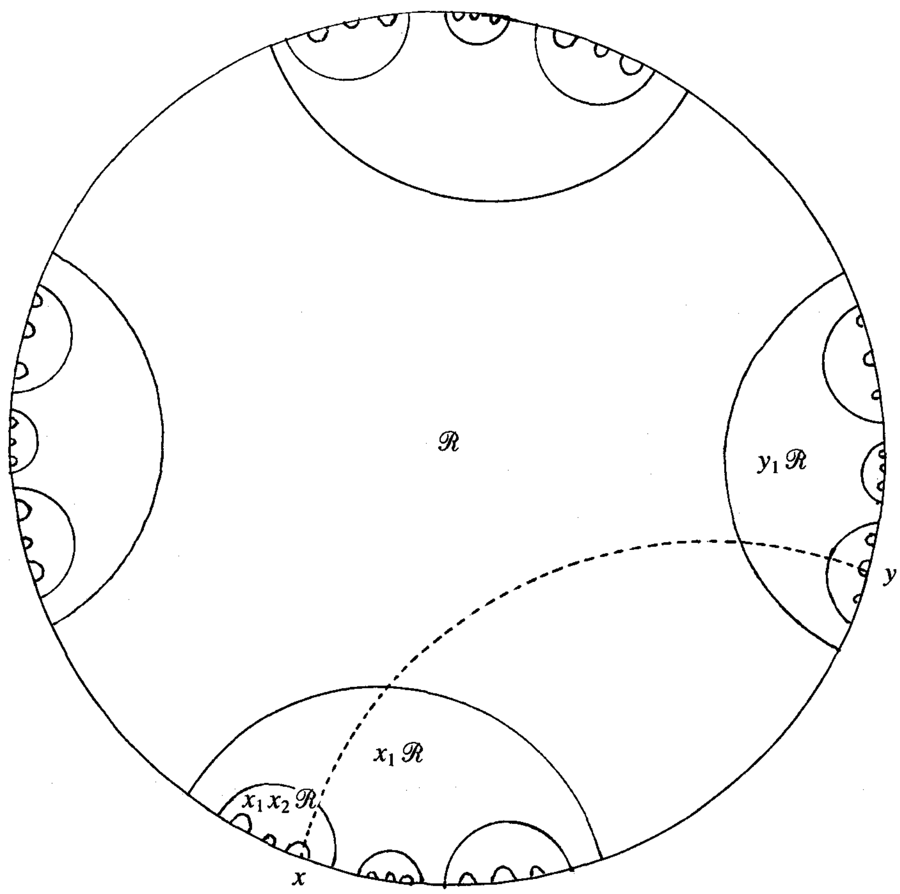

Fig. 2

continuous on $\tilde{\Sigma}$ (see [28] for a similar argument). Observe that $g \circ \sigma^n(\ldots x_2^{-1}x_1^{-1}y_1y_2\ldots)$ is the noneuclidean length of the segment of (x, y) lying in $y_1y_2\ldots y_n\mathcal{R}$ (if $n>0$) or $x_1x_2\ldots x_n\mathcal{R}$ (if $n<0$).

Consider the geodesic flow on the unit tangent bundle $T(H^2/\Gamma)$. The set F of points in $T(H^2/\Gamma)$ whose orbits correspond to geodesics in H^2 with endpoints x, y both in Λ is a closed subset of $T(H^2/\Gamma)$, since Λ is a closed subset of S^1. Call the flow restricted to F the *restricted geodesic flow*. What we have shown is that the restricted geodesic flow has a representation as a suspension flow over a shift of finite type with Hölder continuous height function g (Section 1).

THEOREM 8. *The height function g is cohomologous to the distortion function f.*

Proof. It suffices to show that for every n-periodic sequence

$$\dots x_1 x_2 \dots x_n x_1 x_2 \dots x_n \dots = \tilde{x} \quad \text{in } \tilde{\Sigma},$$

$$S_n f(\tilde{x}) = S_n g(\tilde{x}) \tag{11.2}$$

([1], Theorem 1.28). Let $x = x_1 x_2 \dots x_n x_1 \dots$ and $x' = x_n^{-1} x_{n-1}^{-1} \dots x_1^{-1} x_n^{-1} \dots$, and consider the geodesic (x', x) in H^2. The transformation $x_1 x_2 \dots x_n \in \Gamma$ maps this geodesic onto itself, mapping that segment in $(x_1 x_2 \dots x_{kn+i})\mathcal{R}$ onto the segment in

$$(x_1 x_2 \dots x_n)(x_1 x_2 \dots x_{kn+i})\mathcal{R}.$$

Consequently, the corresponding geodesic on H^2/Γ is periodic, with period $S_n g(\tilde{x})$. It follows that for any point z on (x', x) in H^2, $S_n g(\tilde{x})$ is the length of the segment $(z, (x_1 x_2 \dots x_n) z)$ of (x', x).

Choose z on (x', x) very near x (in the euclidean metric). Near x the geodesic (x', x) looks like a line emanating from x perpendicular to S^1, and the transformation $x_1 x_2 \dots x_n$ looks like a homothety (expansion) with expansion rate

$$\left| \left(\frac{d}{d\xi} (x_1 x_2 \dots x_n)(\xi) \right)_{\xi = x} \right| = e^{S_n f(x)}.$$

Therefore, for z near x the noneuclidean distance from z to $(x_1 x_2 \dots x_n) z$ is approximately

$$\int_{\varepsilon}^{\varepsilon e^{S_n f(x)}} \frac{2dt}{1-(1-t)^2} \approx S_n f(x).$$

Letting $z \to x$, we obtain (11.2). □

Consider again the suspension flow over the shift $(\tilde{\Sigma}, \sigma)$ with height function g. If g is nonlattice then the suspension flow is topologically mixing ([19], Section 5). Conversely, if g is lattice then the suspension flow is not mixing for *any* invariant measure (if g is cohomologous to an integer-valued function then the flow only returns (approximately) to its initial point at (approximately) integer times). It is known [24] that there exists an invariant measure for the restricted geodesic flow that is mixing. (Note: this invariant measure [24] is supported by the subset F of $T(H^2/\Gamma)$ carrying the restricted geodesic flow.) Therefore,

COROLLARY 11.1. *The distortion function f of a Fuchsian Schottky group is nonlattice.*

It follows that the suspension flow on $\bar{\Sigma}_g$ is topologically mixing (Section 5). The periodic orbits of the geodesic flow correspond to those of the suspension flow, and have the same lengths. Consequently, (5.2)–(5.3) translate to corresponding statements about the periodic orbits of the geodesic flow. In particular, if S is the measure on F corresponding to the measure M (cf. (5.1)) on $\bar{\Sigma}_g$, then

COROLLARY 11.2. *For any $\varepsilon > 0$ and any continuous $G: F \to \mathbf{R}$,*

$$\#\{\tau\colon \lambda(\tau) \leq a\} \sim e^{a\delta}/a\delta, \tag{11.3}$$

$$\frac{\#\{\tau\colon \lambda(\tau) \leq a, |\tau(G) - \int G\,dS| > \varepsilon\}}{\#\{\tau\colon \lambda(\tau) \leq a\}} \to 0 \tag{11.4}$$

as $a \to \infty$.

Note. S coincides (up to a scalar factor) with Sullivan's invariant measure [31] for the restricted geodesic flow. (This is because S is an ergodic invariant measure that induces a measure on the endpoints (x, y) at ∞ of geodesics that is equivalent to $\nu_{-\delta f}(dx)\,\nu_{-\delta f}(dy)$.) Thus (11.4) implies that most closed geodesics are approximately distributed on $T(H^2/\Gamma)$ according to the Sullivan measure.

Now let Γ be a nonexceptional Fuchsian group without parabolic elements (Section 10). Then Γ has a Schottky subgroup Γ_0. If the distortion function of Γ were lattice then the distortion function of Γ_0 would also be lattice (this is easily proved by looking at elements of Λ_Γ and Λ_{Γ_0} that have periodic expansions). Therefore,

COROLLARY 11.3. *The distortion function f of a nonexceptional Fuchsian group with no parabolic elements is nonlattice.*

The results about the distortion function we have just obtained may be recast in a form that submerges the role of the shift and the associated symbolic dynamics. Let Γ be any Kleinian group and Λ its limit set. A *cocycle* is a continuous function $U: \Gamma \times \Lambda \to \mathbf{R}$ satisfying

$$U(\gamma_1\gamma_2, x) = U(\gamma_1, \gamma_2 x) + U(\gamma_2, x);$$

a *coboundary* is a cocycle of the form

$$U(\gamma, x) = W(\gamma x) - W(x)$$

for a suitable continuous $W\colon \Lambda \to \mathbf{R}$. Two cocycles are *cohomologous* if their difference is a coboundary; a cocycle is *lattice* if it is cohomologous to a cocycle valued in a discrete subgroup of **R**. Define the *distortion* cocycle U by

$$U(\gamma, x) = \log|\gamma'(x)|.$$

If the distortion cocycle is lattice then the distortion function is also lattice (for Γ of the types considered in Sections 9–10). Therefore,

COROLLARY 11.4. *If Γ contains a Schottky subgroup then the distortion cocyle is nonlattice.*

This is similar to [32], Theorem 6, but seems more general. Sullivan's approach is totally different from ours, and does not seem to generalize easily.

Finally, observe that Theorem 8 and Corollary 11.1 hold also for non-Fuchisan Schottky groups. The proofs are identical except that the geometry must be carried out in H^3 instead of H^2.

12. Distribution of noneuclidean lattice points and fundamental polygons

Let Γ be a nonexceptional Fuchsian group with no parabolic elements, and let $z \in H^2$ ($=D$) be such that z is not a fixed point of any $\gamma \in \Gamma$. Let d_H=hyperbolic distance.

THEOREM 9. *As $a \to \infty$*

$$\#\{\gamma \in \Gamma\colon d_H(0, \gamma z) \leq a\} \sim Ce^{a\delta} \tag{12.1}$$

for a suitable constant $C \in (0, \infty)$. Moreover, if P^a is the uniform probability distribution on $\{\gamma z\colon d_H(0, \gamma z) \leq a,\ \gamma \in \Gamma\}$ then as $a \to \infty$

$$P^a \xrightarrow{D} \text{Patterson measure.} \tag{12.2}$$

Recall that the Patterson measure is the normalized δ-dimensional Hausdorff measure H^δ_Λ on the limit set $\Lambda = \Lambda_\Gamma$. The statement (12.2) means that for any continuous function g on the closed unit disc $\bar{D}$

$$\int g dP^a \to \int g dH^\delta_\Lambda. \tag{12.3}$$

The result (12.1) is a special case of the principal result of [14] (see also [17]), which states that (12.1) holds for *all* finitely generated groups, gives a formula for the constant C, and provides an error estimate. Our method may generalize to other groups, but probably it cannot be adapted to give the precise error estimates of [14]. On the other hand, it seems unlikely that the related Theorems 10–11 below can be obtained by the methods of [14] (or [6], [21]).

For $x=x_1x_2\dots x_n\in\Sigma_*$ $(=\Gamma)$, $x\neq\xi$, define

$$f_*(x)=d_H(0,(x_1x_2\dots x_n)z)-d_H(0,(x_2x_3\dots x_n)z);$$

for $x\in\Sigma$ define $f_*(x)=f(x)$, where f is the distortion function of Γ.

LEMMA 12.1. *$f_*\in\mathcal{F}_\varrho(\Sigma\cup\Sigma_*)$ for some* $0<\varrho<1$.

Proof. Let γ be a linear fractional transformation mapping the unit disc D onto itself. A simple computation shows that

$$\begin{aligned} d_H(0,w)-d_H(0,\gamma w)&=\log\left(\frac{d_E(\gamma w,S^1)}{d_E(w,S^1)}\right)+o(d_E(w,S^1))\\ &=\log|\gamma'(w)|+o(d_E(w,S^1)),\quad w\in D \end{aligned}$$

(d_H, d_E are the hyperbolic and euclidean distances on D and S^1 is the unit circle). Now let $z\in D=H^2$ such that z is not a fixed point of any $\gamma\in\Gamma$; let $x=x_1x_2\dots x_n\in\Sigma_*$, $x'=x_1'x_2'\dots x_m'\in\Sigma_*$, and $x''=x_1''x_2''\dots\in\Sigma=\Lambda$ such that $x_i=x_i'=x_i''$ for $i=1,2,\dots,k$. By (10.4),

$$\begin{aligned} &d_E(xz,x'z)\le C\varrho^k,\\ &d_E(xz,x'')\le C\varrho^k,\\ &d_E(x'z,x'')\le C\varrho^k \end{aligned}$$

(see also Lemma 9.1 for Schottky groups). Hence, for a suitable $0<C'<\infty$,

$$\begin{aligned} &|f_*(x)-f_*(x'')|\le C'\varrho^k,\\ &|f_*(x')-f_*(x'')|\le C'\varrho^k. \end{aligned}\qquad\square$$

Proof of Theorem 9. (12.1) follows immediately from Theorem 4, Corollary 11.3, and Lemma 12.1. In fact, Theorem 4 implies that for any sequence $y=y_1y_2\dots y_n\in\Sigma_*$,

$$\#\{x_1x_2\dots x_m\in\Sigma_*:\ d_H(0,x_1x_2\dots x_mz)\le a,\ x_i=y_i,\ \forall i\le n\}\sim C(y)\,e^{a\delta}\tag{12.4}$$

where

$$C(y)=\left(\frac{\nu_{-\delta f}(y)}{\delta \int f d\mu_{-\delta f}}\right) h_*(\xi),$$

$$\nu_{-\delta f}(y)=\nu_{-\delta f}\{x=x_1x_2\ldots \in \Sigma\colon x_i=y_i,\ \forall i\leqslant n\}.$$

Now the cylinder sets $\{x\in\Sigma\colon x_i=y_i,\ \forall i\leqslant n\}$ generate the Borel σ-algebra on $\Sigma\cong\Lambda$, and $\nu_{-\delta f}$ is the Patterson measure H_Λ^δ on Λ. Consequently, (12.4) implies that for any angular sector

$$A=\{re^{i\theta}\colon 1-\varepsilon\leqslant r\leqslant 1,\ \theta_1\leqslant\theta\leqslant\theta_2\},$$

$$\lim_{a\to\infty} P^a(A)=H_\Lambda^\delta(A).$$

The result (12.3), hence (12.2), follows by a routine approximation argument. □

Now let Γ be a Schottky group and let z by any point of discontinuity, i.e., $z\in\mathbf{C}\cup\{\infty\}-\Lambda$.

THEOREM 10. *As* $\varepsilon\to 0$

$$\#\{\gamma\in\Gamma\colon d_E(\gamma z,\Lambda)>\varepsilon\}\sim C\varepsilon^{-\delta}$$

for a suitable constant $C\in(0,\infty)$. *If* P^ε *is the uniform probability distribution on* $\{\gamma z\colon d_E(\gamma z,\Lambda)>\varepsilon,\ \gamma\in\Gamma\}$ *then as* $\varepsilon\to\infty$

$$P^\varepsilon \xrightarrow{D} \text{Patterson measure.}$$

There is a similar result for nonexceptional Fuchsian groups without parabolic elements and $z\in S^1-\Lambda$. Undoubtedly it holds in even greater generality.

The proof of Theorem 10 is virtually the same as that of Theorem 9; this time we use the function

$$f_*(x)=\log\left(\frac{d_E((x_2x_3\ldots x_n)z,\Lambda)}{d_E((x_1x_2\ldots x_n)z,\Lambda)}\right)$$

for $x=x_1x_2\ldots x_n\in\Sigma_*$, $f_*(x)=f(x)$ for $x\in\Sigma$. In proving the analogue of Lemma 12.1 one must use (9.5) in place of (10.4).

Finally we consider the polygons in a noneuclidean tessellation. For simplicity we consider only Schottky groups. Let $\mathscr{R}$ be the natural fundamental region, i.e., that part of $\mathbf{C}\cup\{\infty\}$ exterior to each of the circles $Q_1, Q_{-1}, \ldots, Q_{-k}$.

THEOREM 11. *As* $\varepsilon\to\infty$

$$\#\{\gamma\in\Gamma\colon \operatorname{Area}(\gamma\mathcal{R})>\varepsilon\}\sim C\varepsilon^{-\delta/2} \tag{12.5}$$

for a suitable constant $C\in(0,\infty)$.

Note. Area means euclidean area. It doesn't matter whether the euclidean area for the sphere $S^2=\mathbf{C}\cup\{\infty\}$ or the plane $\mathbf{C}$ is used.

Proof. For $x=x_1x_2\dots x_n\in\Sigma_*$, $x\neq\xi$, define

$$f_*(x)=\frac{1}{2}\log\left(\frac{\operatorname{Area}(x_2x_3\dots x_n\mathcal{R})}{\operatorname{Area}(x_1x_2\dots x_n\mathcal{R})}\right)$$

and for $x\in\Sigma$ define $f_*(x)=f(x)=$distortion function. We must show that $f_*\in\mathcal{F}_\varrho(\Sigma\cup\Sigma_*)$ for some $\varrho\in(0,1)$; (12.5) will then follow from Theorem 4 and Corollary 11.1.

Let $x=x_1x_2\dots\in\Sigma=\Lambda$, $x'=x_1'x_2'\dots x_n'\in\Sigma_*=\Gamma$, $x''=x_1''x_2''\dots x_m''\in\Sigma_*$ such that $x_i=x_i'=x_i''$ for $i=1,2,\dots,k$. Then each of $x, x'\mathcal{R}, x''\mathcal{R}$ is contained in the disc $D_{x_1x_2\dots x_k}$ (see (9.4)) which by Lemma 9.1 has diameter $\leqslant C\varrho^k$. Now

$$\operatorname{Area}(x_1x_2\dots x_n\mathcal{R})=\iint_{x_2x_3\dots x_n\mathcal{R}}\left|\frac{d}{dz}x_1(z)\right|^2dm(z),$$

$$\operatorname{Area}(x_1'x_2'\dots x_m'\mathcal{R})=\iint_{x_2'x_3'\dots x_m'\mathcal{R}}\left|\frac{d}{dz}x_1'(z)\right|^2dm(z),$$

where m denotes Lebesgue measure. The diameters of $x_2x_3\dots x_n\mathcal{R}$ and $x_2'x_3'\dots x_m'\mathcal{R}$ are $\leqslant C\varrho^{k-1}$, and $(d/dz)x_1$ is Lipschitz on $\bar{D}_{x_2x_3\dots x_k}$ (since $\bar{D}_{x_2x_3\dots x_k}$ lies outside the isometric circle of x_1; see the proof of Lemmas 9.1–9.2). Consequently,

$$|f_*(x)-f_*(x')|\leqslant C'\varrho^k,$$

$$|f_*(x)-f_*(x'')|\leqslant C'\varrho^k$$

for some $C'<\infty$. This proves that $f_*\in\mathcal{F}_\varrho(\Sigma\cup\Sigma_*)$. □

13. Packing and covering functions of the limit set

In this section we shall describe a more elaborate application of the renewal theory, this to the geometry of the limit set Λ. Let F be a finite set of points in the plane. Say that F is an *ε-covering* of Λ if every point of Λ is within ε of some point of F; say that F is an *ε-packing* of Λ if $F\subset\Lambda$ and no two points of F are within ε of each other. Define the *covering function* $N(\varepsilon)$ of Λ to be the minimum cardinality of an ε-covering of Λ, and define the *packing function* $M(\varepsilon)$ of Λ to be the maximum cardinality of an ε-packing of Λ. (Note: In this section the only distance function considered is the euclidean distance on $\mathbf{R}^2=\mathbf{C}$.)

THEOREM 12. *If Λ is the limit set of a Schottky group then as $\varepsilon\to 0$*

$$N(\varepsilon)\sim C\varepsilon^{-\delta}, \tag{13.1}$$

$$M(\varepsilon)\sim C'\varepsilon^{-\delta} \tag{13.2}$$

for suitable constants $C, C'\in(0,\infty)$.

The functions $N(\varepsilon)$ and $M(\varepsilon)$ are used to define dimensional quantities usually called the *metric entropy* and *capacity* (better terminology might be *covering dimension* and *packing dimension*):

$$\text{metric entropy}=\lim_{\varepsilon\to 0}\left(\frac{\log N(\varepsilon)}{\log\varepsilon^{-1}}\right),$$

$$\text{capacity}=\lim_{\varepsilon\to 0}\left(\frac{\log M(\varepsilon)}{\log\varepsilon^{-1}}\right)$$

provided these limits exist [9]. Theorem 12 shows that the Hausdorff dimension, metric entropy, and capacity of Λ are all equal. This is apparently not true for limit sets of discrete groups with parabolic elements [33].

It is natural to ask about the distribution of points in an economical ε-covering (or an ε-packing). For $\varepsilon>0$ let F_ε be an ε-covering of Λ of minimum cardinality and let P^ε be the uniform probability measure on F_ε.

THEOREM 13. *If Λ is the limit set of a Schottky group then as $\varepsilon\to 0$*

$$P^\varepsilon\xrightarrow{D}\text{Patterson measure.} \tag{13.3}$$

In other words, if $g:\mathbf{C}\to\mathbf{R}$ is any bounded, continuous function then as $\varepsilon\to 0$

$$\int g\,dP^\varepsilon\to\int g\,dH_\Lambda^\delta$$

where H^{δ}_{Λ} is the Patterson measure (normalized δ-dimensional Hausdorff measure on Λ). There is a similar theorem for maximal ε-packings.

The key to Theorems 12–13 is the approximate local self-similarity of Λ. Let J be the intersection of Λ with a small disc, and let $F\colon \Lambda\to\Lambda$ be the Nielsen map (Sections 9, 10). Since F is conformal, F acts on J approximately as a homothety; consequently, any ε-covering of J is mapped by F to an $\varepsilon\varrho$-covering of $F(J)$ for a suitable $\varrho>0$. (Note: for the limit set of a semigroup of contractive homotheties the analogues of Theorems 12–13 are simpler; see [13].)

In proving Theorems 12–13 we shall, for ease of exposition, only discuss (13.1).

We shall begin by modifying the symbolic dynamics constructed in Section 9. Let Γ be a Schottky group generated by $T_1, T_2, \ldots, T_k$, as in Section 9; recall that Γ may be identified with the set Σ_* of finite admissible sequences from $\Gamma_1=\{T_1, T_1^{-1}, T_2, \ldots, T_k^{-1}\}$, and that the limit set Λ may be identified with the set Σ of infinite admissible sequences. Let Γ_r, $r\geqslant 1$, be the set of all sequences $x_1x_2\ldots x_r\in\Sigma_*$ of length r, and let Σ^r (Σ^r_*) be the set of infinite (finite) admissible sequences from Γ_r. (Admissible transitions are defined as follows: if

$$y=x_1x_2\ldots x_r\in\Gamma_r,$$

$$y'=x_1'x_2'\ldots x_r'\in\Gamma_r,$$

then yy' is admissible iff $x_rx_1'\neq$identity.) There is a natural homeomorphism $\Sigma\to\Sigma^r$ given by

$$x_1x_2\ldots\to(x_1x_2\ldots x_r)(x_{r+1}x_{r+2}\ldots x_{2r})\ldots. \tag{13.4}$$

Thus, for each $r\geqslant 1$ the limit set Λ may be identified with Σ^r, so there are infinitely many choices of "symbolic dynamics".

Recall from the proof of Lemmas 9.1–9.2 that there exist $0<\alpha<1$ and $r\geqslant 1$ such that for each $x_1x_2\ldots x_r\in\Gamma_r$ the derivative of the function $(x_1x_2\ldots x_r)$ in absolute value is $\leqslant\alpha$ on $\bigcup_{\gamma\neq x_r^{-1}}\bar D_\gamma$. Fix such an r; henceforth we will only use the sequence spaces Σ^r, Σ^r_*. For notational convenience we will drop the superscript r and refer to these spaces as Σ, Σ_*. The shift on $\Sigma\cup\Sigma_*$ will be denoted by σ_*. All entries x_i, y_i, z_i, w_i occurring in sequences, either finite or infinite, are henceforth from the alphabet Γ_r.

(Note: The space Σ_* of finite sequences from Γ_r can no longer be identified with Γ, or even a subgroup of Γ. This will not matter.)

Consider now the shift $\sigma\colon\Sigma\to\Sigma$; the corresponding map on Λ is $F^{(r)}$, where F is the Nielsen map defined in Section 9. Define

$$\bar{f}^{(r)}(z) = \log|(F^{(r)})'(z)|, \quad z \in \Lambda,$$

and let $f: \Sigma \to \mathbf{R}$ be the pullback of $\bar{f}^{(r)}$ to Σ. Because of our choice of r,

$$f(x) > 0, \quad \forall x \in \Sigma \tag{13.5}$$

(see the proof of Lemma 9.1).

LEMMA 13.1. *The function f is nonlattice, and the unique $\delta > 0$ such that $\lambda_{-\delta f} = 1$ (cf. (2.3)) is the same δ as for the distortion function of Γ.*

Proof. For this proof let g denote the distortion function of Γ (cf. Section 9). By construction, the pullback of f to the original sequence space (via the map (13.4)) coincides with $S_r g$, and the Perron-Frobenius operator $\mathscr{L}_{zf}$ pulls back to $\mathscr{L}^r_{zg}$. Consequently $\lambda_{zf} = \lambda^r_{zg}$. The statement about δ follows immediately. That f is nonlattice follows from the fact that g is nonlattice (Corollary 11.1) and Theorem B (Section 7), together with the spectral radius formula. □

Let $D_{x_1 x_2 \dots x_n}$, $x_1 x_2 \dots x_n \in \Sigma_*$, be the discs defined by (9.1). (This is a subset of the system of disks defined by (9.1), because $x_i \in \Gamma_r$; hence, (9.2)–(9.5) still hold.) Define functions f_1, f_2: $\Sigma \cup \Sigma_* \to \mathbf{R}$ by

$$f_i(x) = f(x), \quad x \in \Sigma,$$

$$f_1(x_1 x_2 \dots x_n) = \inf\left\{\log\left|\frac{d}{dz} x_1^{-1}(z)\right| : z \in \bar{D}_{x_1 x_2 \dots x_n}\right\},$$

$$f_2(x_1 x_2 \dots x_n) = \sup\left\{\log\left|\frac{d}{dz} x_1^{-1}(z)\right| : z = \bar{D}_{x_1 x_2 \dots x_n}\right\},$$

$$f_i(\xi) = 0.$$

LEMMA 13.2. *The functions f_1 and f_2 satisfy*

$$f_1, f_2 \in \mathscr{F}_\varrho(\Sigma \cup \Sigma_*) \tag{13.6}$$

$$0 < f_1(x) \leqslant f_2(x), \quad \forall x \in \Sigma \cup \Sigma_*, \ x \neq \xi, \tag{13.7}$$

$$f_1(x_1 x_2 \dots x_n) \leqslant f_1(x_1 x_2 \dots x_n x_{n+1} \dots x_{n+j}), \quad \forall x_1 x_2 \dots x_{n+j} \in \Sigma_*, \tag{13.8}$$

$$f_2(x_1 x_2 \dots x_n) \geqslant f_2(x_1 x_2 \dots x_n x_{n+1} \dots x_{n+j}), \quad \forall x_1 x_2 \dots x_{n+j} \in \Sigma_*. \tag{13.9}$$

Proof. (13.8)–(13.9) follows from (9.2). The positivity of f_1 follows from our choice of r. The Hölder continuity of f_1, f_2 follows from (9.5). □

Define

$$\Sigma_m = \{x \in \Sigma_*: x = x_1 x_2 \dots x_m\}, \quad m = 1, 2, \dots, \tag{13.10}$$

i.e., Σ_m is the set of all sequences of length m with allowable transitions. Define

$$\alpha_m = \sup_{n \geq 0} \max\{S_n f_2(x) - S_n f_1(x): x \in \Sigma_{n+m}\}. \tag{13.11}$$

Observe that $\alpha_m \leq \Sigma_m^\infty(\mathrm{var}_j f_2 + \mathrm{var}_j f_1)$, so by (13.6)

$$\alpha_m \to 0 \quad \text{as} \quad m \to \infty. \tag{13.12}$$

LEMMA 13.3. *Let*

$$x = x_1 x_2 \dots x_{n+m} \in \Sigma_{n+m} \quad \textit{where} \quad n, m \geq 1.$$

For any $z_1, z_2 \in \bar{D}_{x_{n+1} x_{n+2} \dots x_{n+m}}$,

$$e^{S_n f_1(x)} \leq \frac{\mathrm{dist}(z_1, z_2)}{\mathrm{dist}((x_1 x_2 \dots x_n) z_1, (x_1 x_2 \dots x_n) z_2)} \leq e^{S_n f_2(x)}.$$

Here $x_1 x_2 \dots x_n$ denotes both a sequence in Σ_* and an element of the group Γ. To prove Lemma 13.3 it suffices, by induction, to consider the case $n=1$; for $n=1$, the result follows from elementary calculus.

LEMMA 13.4. *For any* $x = x_1 x_2 \dots x_{n+m} \in \Sigma_{n+m}$, $n, m \geq 1$,

$$e^{-S_n f_2(x)} \mathrm{diam}(D_{x_{n+1} \dots x_{n+m}}) \leq \mathrm{diam}(D_{x_1 x_2 \dots x_{n+m}}) \leq e^{-S_n f_1(x)} \mathrm{diam}(D_{x_{n+1} \dots x_{n+m}}).$$

This is a straightforward consequence of Lemma 13.3.

LEMMA 13.5. *There exists $C < \infty$ with the following property: if $\mathscr{D}$ is any collection of pairwise disjoint discs from the set $\{D_{x_1 x_2 \dots x_n}: x_1 x_2 \dots x_n \in \Sigma_*\}$ such that each disc in $\mathscr{D}$ has diameter $\geq C\varepsilon$, then the distance between any two distinct discs in $\mathscr{D}$ is $> \varepsilon$.*

Note. The constant C does not depend on ε.

Proof. Consider a pair of disjoint discs whose distance is $\leq \varepsilon$. Then by (9.2)–(9.3) these discs are contained in discs $D_{x_1x_2\dots x_nx_{n+1}}$, $D_{x_1x_2\dots x_nx'_{n+1}}$ whose distance is $\leq\varepsilon$. It suffices to show that $\operatorname{diam} D_{x_1x_2\dots x_{n+1}}<C\varepsilon$.

If $\operatorname{dist}(D_{x_1x_2\dots x_nx_{n+1}}, D_{x_1x_2\dots x_nx'_{n+1}})\leq\varepsilon$ then by Lemma 13.3

$$e^{S_{n-1}f_2(x_1x_2\dots x_n)}\varepsilon\geq\operatorname{dist}(D_{x_nx_{n+1}}, D_{x_nx'_{n+1}})$$

On other hand, Lemma 13.4 implies that

$$\operatorname{diam} D_{x_1x_2\dots x_{n+1}}\leq e^{-S_nf_1(x_1x_2\dots x_{n+1})}\operatorname{diam} D_{x_{n+1}}.$$

Consequently, since there are only finitely many $D_{x_{n+1}}$ and $D_{x_nx_{n+1}}$, it suffices to show that there is a constant $C'<\infty$ such that

$$S_{n-1}f_2(x)-S_nf_1(x)\leq C'$$

for every sequence $x\in\Sigma_*$. But this follows from (13.11)–(13.12). □

Let K be any compact subset of $\mathbf{R}^2$; define $N(\varepsilon, K)$ to be the minimum cardinality of an ε-covering of K (note: $N(\varepsilon)=N(\varepsilon,\Lambda)$). Clearly, $N(\varepsilon, K)$ is a nonincreasing, integer-valued, right-continuous function of $\varepsilon>0$. If K is the union of pairwise disjoint, compact sets K_i such that $\operatorname{dist}(K_i, K_j)>2\varepsilon$ if $i\neq j$ then $N(\varepsilon, K)=\Sigma_iN(\varepsilon, K_i)$.

Define subsets $\mathscr{D}_\varepsilon^m$ of Σ_* as follows. Fix $\gamma_m\geq0$ real and $m\geq1$ integer. Let $\mathscr{D}_\varepsilon^m$ be the set of all $x_1x_2\dots x_{n+m}\in\Sigma_*$ such that

$$S_nf_1(x_1x_2\dots x_{n+m})>(\log\varepsilon^{-1})-\gamma_m, \tag{13.13}$$

$$S_jf_1(x_1x_2\dots x_{j+m})\leq(\log\varepsilon^{-1})-\gamma_m,\quad \forall j<n. \tag{13.14}$$

Lemmas 13.4–13.5 and relation (13.12) imply that if γ_m is chosen sufficiently *large* then for each pair of distinct sequences $x_1x_2\dots x_{n+m}, x'_1x'_2\dots x'_{n+m}\in\mathscr{D}_\varepsilon^m$,

$$\operatorname{dist}(\bar{D}_{x_1x_2\dots x_{n+m}}, \bar{D}_{x'_1x'_2\dots x'_{n+m}})>2\varepsilon. \tag{13.15}$$

Let $x=x_1x_2\dots\in\Sigma$; since $S_nf(x)\to\infty$ as $n\to\infty$ and f_1 is a Hölder continuous extension of f, there exists $n<\infty$ such that (13.13) holds. Consequently,

$$\Lambda=\bigcup_{\mathscr{D}_\varepsilon^m}\Lambda_{x_1x_2\dots x_{n+m}}, \tag{13.16}$$

where

$$\Lambda_{x_1x_2\dots x_n}=\Lambda\cap\bar{D}_{x_1x_2\dots x_n}.$$

It follows from (13.15)–(13.16) that

$$N(\varepsilon)=N(\varepsilon,\Lambda)=\sum_{\mathcal{D}_\varepsilon^m} N(\varepsilon,\Lambda_{x_1x_2\dots x_{n+m}}).$$

Fix $w=w_1w_2\dots w_k\in\Sigma_k$, some $k\geqslant 1$, and let $g_w(x)=1$ if $x_i=w_i$ for $i=1,2,\dots,k$, $g_w(x)=0$ otherwise. As $\varepsilon\to 0$ the sequences in $\mathcal{D}_\varepsilon^m$ become longer; in particular, for all ε sufficiently small each sequence in $\mathcal{D}_\varepsilon^m$ has length $\geqslant k$. By the nesting property (9.2)–(9.3).

$$\begin{gathered}\Lambda_{w_1w_2\dots w_k}=\bigcup_{x\in\mathcal{D}_\varepsilon^m:\,g_w(x)=1}\Lambda_{x_1x_2\dots x_{n+m}} \quad\Rightarrow\\ N(\varepsilon,\Lambda_{w_1w_2\dots w_k})=\sum_{\mathcal{D}_\varepsilon^m} g_w(x)\,N(\varepsilon,\Lambda_{x_1x_2\dots x_{n+m}}).\end{gathered}\tag{13.17}$$

Let $x=x_1x_2\dots x_{n+m}\in\mathcal{D}_\varepsilon^m$. *If* F is an ε-covering of $\Lambda_{x_1x_2\dots x_{n+m}}$ then by Lemma 13.3, $(x_1x_2\dots x_n)^{-1}F$ is an $\varepsilon\exp(S_nf_2(x))$-covering of $\Lambda_{x_{n+1}\dots x_{n+m}}$. Similarly, if G is an $\varepsilon\exp(S_nf_1(x))$-covering of $\Lambda_{x_{n+1}\dots x_{n+m}}$ then $(x_1x_2\dots x_n)G$ is an ε-covering of $\Lambda_{x_1x_2\dots x_{n+m}}$. Hence, by (13.17) and (13.11), for all $\varepsilon>0$ sufficiently small,

$$\begin{aligned}\sum_{y\in\Sigma_m}\sum_{x\in\mathcal{D}_\varepsilon(y)} g_w(x)\,N(\varepsilon e^{S_nf_1(x)+\alpha_m},\Lambda_{y_1y_2\dots y_m})&\leqslant N(\varepsilon,\Lambda_{w_1w_2\dots w_k})\\ &\leqslant\sum_{y\in\Sigma_m}\sum_{x\in\mathcal{D}_\varepsilon(y)} g_w(x)\,N(\varepsilon e^{S_nf_1(x)},\Lambda_{y_1y_2\dots y_m})\end{aligned}\tag{13.18}$$

where Σ_m is defined by (13.10) and, for each $y\in\Sigma_m$,

$$\mathcal{D}_\varepsilon(y)=\{x\in\mathcal{D}_\varepsilon^m:\ x_{n+i}=y_i,\ \forall i=1,2,\dots,m\}.$$

The sums in (13.18) are nearly of the same form as the sums considered in Corollary 3.2. Observe that by (13.8), if (13.13) holds then (13.14) is implied by

$$S_{n-1}f_1(x_1x_2\dots x_{n+m})\leqslant(\log\varepsilon^{-1})-\gamma_m;\tag{13.19}$$

on the other hand, if (13.14) holds then

$$S_{n-1}f_1(x_1x_2\dots x_{n+m})\leqslant(\log\varepsilon^{-1})-\gamma_m+\alpha_m.\tag{13.20}$$

For $y \in \Sigma_m$, define

$$\mathscr{D}_\varepsilon'(y) = \{x = x_1 x_2 \dots x_{n+m} \in \Sigma_*: x_{n+i} = y_i, \ \forall i \leq m \text{ and (13.13), (13.19) hold}\},$$

$$\mathscr{D}_\varepsilon''(y) = \mathscr{D}_\varepsilon(y) - \mathscr{D}_\varepsilon'(y).$$

Then (13.18) implies that for all $\varepsilon > 0$ sufficiently small,

$$\begin{aligned}\sum_{y \in \Sigma_m} \sum_{x \in \mathscr{D}_\varepsilon'(y)} g_w(x) N(\varepsilon e^{S_n f_1(x) + \alpha_m}, \Lambda_{y_1 y_2 \dots y_m}) &\leq N(\varepsilon, \Lambda_{w_1 w_2 \dots w_k}) \\ &\leq \sum_{y \in \Sigma_m} \sum_{x \in \mathscr{D}_\varepsilon'(y)} g_w(x) N(\varepsilon e^{S_n f_1(x)}, \Lambda_{y_1 y_2 \dots y_m}) \\ &\quad + \sum_{y \in \Sigma_m} \sum_{x \in \mathscr{D}_\varepsilon''(y)} g_w(x) N(e^{-\gamma_m}, \Lambda_{y_1 y_2 \dots y_m}).\end{aligned} \tag{13.21}$$

Now each of the sums $\Sigma_{x \in \mathscr{D}_\varepsilon'(y)}$ is of the form $\Sigma_{\tilde{y}: \sigma_* \tilde{y} = y} N_G((\log \varepsilon^{-1}) - \gamma_m, \tilde{y})$ for a suitable monotone G, where $N_G(a, y)$ is defined by (3.10), hence Corollary 3.2 applies to each. The next order of business is to show that if m is chosen very large (thus α_m is small, by (13.12)) then the ratio of the right and left sides of (13.21) is close to 1 as $\varepsilon \to 0$.

LEMMA 13.6. *There exists $C < \infty$ such that for any $m \geq 1$, $y \in \Sigma_m$, $0 < \varepsilon \leq \varepsilon(m)$,*

$$\# \mathscr{D}_\varepsilon''(y) \leq \varepsilon^{-\delta}(C e^{-\delta \gamma_m} \alpha_m) \tag{13.22}$$

Note. $\#F$=cardinality of F.

Proof. By (13.19)–(13.20), if $x \in \mathscr{D}_\varepsilon''(y)$ then

$$(\log \varepsilon^{-1}) - \gamma_m \leq S_{n-1} f_1(x) \leq (\log \varepsilon^{-1}) - \gamma_m + \alpha_m.$$

Consequently,

$$\begin{aligned}\# \mathscr{D}_\varepsilon''(y) &\leq \sum_{n=0}^{\infty} \sum_{\tilde{y}: \sigma_* \tilde{y} = y} \sum_{x: \sigma_*^n x = \tilde{y}} 1\{0 \leq S_n f_1(x) + \log \varepsilon + \gamma_m \leq \alpha_m\} \\ &\sim \sum_{\tilde{y}: \sigma_* \tilde{y} = y} C_*(\tilde{y}) \varepsilon^{-\delta} e^{-\delta \gamma_m}(e^{\delta \alpha_m} - 1),\end{aligned}$$

by Theorem 4, where $C_*(\tilde{y})$ is given by (3.4). Since $\gamma_m \geq 0$ and $C_*(\tilde{y})$ is bounded for $y \in \Sigma_*$, (13.22) follows. □

LEMMA 13.7. *There exists $C<\infty$ such that for each $w=w_1w_2\dots w_k\in\Sigma_*$,*

$$\frac{N(\varepsilon_1, \Lambda_{w_1w_2\dots w_k})}{N(\varepsilon_2, \Lambda_{w_1w_2\dots w_k})}\leq C(\varepsilon_1/\varepsilon_2)^{-\delta} \tag{13.23}$$

for all $0<\varepsilon_1\leq\varepsilon_2\leq\varepsilon(w)$.

Proof. Recall that $N(\varepsilon, K)\geq 1$ for $\varepsilon>0$ and $K\neq\emptyset$ and that $N(\varepsilon, K)$ is nonincreasing in ε. Choose $\varepsilon(w)>0$ such that (13.21) holds for all $0<\varepsilon<\varepsilon(w)$. Then for $\varepsilon<\varepsilon(w)$

$$\sum_{y\in\Sigma_m}\sum_{x\in\mathscr{D}'_\varepsilon(y)} g_w(x)\leq N(\varepsilon, \Lambda_{w_1w_2\dots w_k})\leq\bar{N}\left\{\sum_{y\in\Sigma_m}\left\{\sum_{x\in\mathscr{D}'_\varepsilon(y)} g_w(x)+\sum_{x\in\mathscr{D}''_\varepsilon(y)} g_w(x)\right\}\right\} \tag{13.24}$$

where

$$\bar{N}=\max\{N(e^{-\gamma_m}, \Lambda_{y_1y_2\dots y_m})\colon y_1y_2\dots y_m\in\Sigma_m\}.$$

It follows from Corollary 3.1 that

$$\sum_{x\in\mathscr{D}'_\varepsilon(y)} g_w(x)\sim C_y\varepsilon^{-\delta} \tag{13.25}$$

for some $0<C_y<\infty$. Inequality (13.23) follows from (13.22), (13.24), (13.25), and (13.5). □

Consider again the double inequality (13.21). Fix $y\in\Sigma_m$; define

$$G_y(t)=N(e^{t-\gamma_m}, \Lambda_{y_1y_2\dots y_m}),$$

$$\tilde{G}_y(t)=N(e^{t-\gamma_m+\alpha_m}, \Lambda_{y_1y_2\dots y_m})=G_y(t+\alpha_m).$$

Then

$$\sum_{x\in\mathscr{D}'_\varepsilon(y)} g_w(x)\,N(\varepsilon e^{S_nf_1(x)}, \Lambda_{y_1y_2\dots y_m})=\sum_{\tilde{y}:\sigma_*\tilde{y}=y} N_{G_{\tilde{y}}}((\log\varepsilon^{-1})-\gamma_m, \tilde{y}), \tag{13.26}$$

$$\sum_{x\in\mathscr{D}'_\varepsilon(y)} g_w(x)\,N(\varepsilon e^{S_nf_1(x)+\alpha_m}, \Lambda_{y_1y_2\dots y_m})=\sum_{\tilde{y}:\sigma_*\tilde{y}=y} N_{\tilde{G}_{\tilde{y}}}((\log\varepsilon^{-1})-\gamma_m, \tilde{y}), \tag{13.27}$$

where $N_G, N_{\tilde{G}}$ are defined by (3.10). By Corollary 3.2, as $\varepsilon\to 0$

$$N_{G_y}((\log \varepsilon^{-1})-\gamma_m, y) \sim \varepsilon^{-\delta} e^{-\delta\gamma_m} \int_0^\infty G(t)\, F(y, dt), \tag{13.28}$$

$$N_{\tilde{G}_y}((\log \varepsilon^{-1})-\gamma_m, y) \sim \varepsilon^{-\delta} e^{-\delta\gamma_m} \int_0^\infty \tilde{G}(t)\, F(y, dt), \tag{13.29}$$

where $F(y, t)$ is defined by (3.9).

Recall that for each y the measure $F(y, dt)$ is supported by $[0, \|f_1\|_\infty]$, that $\int_0^\infty F(y, dt)$ is bounded above, and that for a suitable $C>0$,

$$F(y, t)-F(y, 0) \geq C(e^{\delta t}-1), \quad \forall y \in \Sigma_* \cup \Sigma$$

(cf. (3.9)). Lemma 13.7 implies that there exists a constant $C'>0$ independent of $m \geq 1$ and $y \in \Sigma_m$ such that

$$G_y(t)/G_y(0) \geq C', \quad \forall\, 0 \leq t \leq \|f_1\|_\infty,$$

$$\tilde{G}_y(t)/G_y(0) \geq C', \quad \forall\, 0 \leq t \leq \|f_1\|_\infty.$$

Consequently, there exists a constant $C''>0$ such that for each $m \geq 1$ and each $y \in \Sigma_m$

$$\int_0^\infty (G_y(t)/G_y(0))\, F(y, dt) \geq C'', \tag{13.30}$$

$$\int_0^\infty (\tilde{G}_y(t)/G_y(0))\, F(y, dt) \geq C''; \tag{13.31}$$

since $G_y(t)/G_y(0)$ and $\tilde{G}_y(t)/G_y(0)$ are ≤ 1 for $t \geq 0$, it therefore follows from Corollary 3.3 that as $m \to \infty$ (recall $\alpha_m \to 0$ as $m \to \infty$)

$$\frac{\int_0^\infty G_y(t)\, F(y, dt)}{\int_0^\infty \tilde{G}_y(t)\, F(y, dt)} \to 1 \tag{13.32}$$

uniformly for $y \in \Sigma_m$.

Now consider the term

$$\sum_{x \in \mathscr{D}''_\varepsilon(y)} g_w(x)\, N(\varepsilon^{-\gamma_m}, \Lambda_{y_1 \ldots y_m}) \tag{13.33}$$

in (13.21). This is bounded above by $(\# \mathscr{D}''_\varepsilon(y))\, G_y(0)$. Hence, by (13.30) and Lemma 13.6, for large m (small α_m) the ratio of (13.33) to

$$\varepsilon^{-\delta} e^{-\delta \gamma_m} \int_0^\infty G_y(t)\, F(y, dt)$$

is small as $\varepsilon \to 0$, uniformly for $y \in \Sigma_m$. Combining this with (13.32) and (13.26)–(13.29), we find that for large m the ratio of the right and left sides of (13.21) is close to 1 for all small $\varepsilon > 0$. Letting $m \to \infty$ we obtain

$$N(\varepsilon, \Lambda_{w_1 w_2 \dots w_k}) \sim C(w)\, \varepsilon^{-\delta} \tag{13.34}$$

as $\varepsilon \to 0$.

Examination of (13.28)–(13.29) shows that the sequence $w = w_1 w_2 \dots w_k$ enters into the asymptotic formula (13.34) only by way of the distribution functions $F(y, t)$. It therefore follows from (3.9) and (3.4) that

$$C(w) = C \int g_w \, d\nu_{-\delta f}$$

for some $0 < C < \infty$ independent of w. Hence, (13.34) implies that as $\varepsilon \to 0$

$$N(\varepsilon) = \sum_{w_1} N(\varepsilon, \Lambda_{w_1}) \sim C \varepsilon^{-\delta}$$

and

$$\frac{N(\varepsilon, \Lambda_{w_1 w_2 \dots w_k})}{N(\varepsilon)} \to \int g_w \, d\nu_{-\delta f} = H_\Lambda^\delta(\Lambda_{w_1 \dots w_k}).$$

This proves Theorems 12–13. □

14. Random walk and Hausdorff measure

Recall that the Patterson measure (normalized δ-dimensional Hausdorff measure on Λ) is the probability measure induced by the measure $\nu_{-\delta f}$ on Σ. The Gibbs measure $\mu_{-\delta f}$ is equivalent to $\nu_{-\delta f}$, hence it induces a probability measure $\xi(dx)$ on Λ equivalent to the Patterson measure.

Let $\{p_\gamma : \gamma \in \Gamma\}$ be a probability distribution on Γ, i.e., $p_\gamma \geq 0$, $\forall \gamma \in \Gamma$ and $\Sigma_\Gamma p_\gamma = 1$. Let $X_1, X_2, \dots$ be a sequence of independent, identically distributed random variables with distribution $\{p_\gamma\}$, i.e. $X_1, X_2, \dots$ are measurable Γ-valued functions on some normalized measure space $(\Omega, \mathscr{F}, P)$ such that

$$P\{X_1=\gamma_1;\ X_2=\gamma_2;\ \ldots;\ X_n=\gamma_n\}=p_{\gamma_1}p_{\gamma_2}\cdots p_{\gamma_n}.$$

For any discrete group Γ of the types considered in Sections 9–10 and any $z\in\mathbf{C}-\Lambda$,

$$\lim_{n\to\infty} X_1X_2\ldots X_n z$$

exists and is independent of z, provided Γ is the smallest group that supports $\{p_\gamma\}$. Define the exit measure $\eta(dx)$ on Λ by

$$\eta(A)=P\{\lim_{n\to\infty} X_1X_2\ldots X_n z\in A\}$$

for Borel measurable $A\subset\Lambda$.

THEOREM 14. *There exists a probability distribution* $\{p_\gamma\}$ *on* Γ *whose exit measure* $\eta=\xi$.

This follows from the construction in [12], which shows that *any* Gibbs measure has a representation as an infinite concatenation of independent, identically distributed random words of finite length from the alphabet $\{1,2,\ldots l\}$.

The probability distribution $\{p_\gamma\}$ given by Theorem 14 is not unique. It would be interesting to have a concrete example of such a distribution. (For self-similar fractals the Hausdorff measure has a very simple and natural representation as an exit measure: cf. [7].)

References

[1] BOWEN, R., *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*. Springer Lecture Notes in Mathematics, 470 (1975).

[2] — Hausdorff dimension of quasicircles. *Inst. Hautes Études Sci. Publ. Math.*, 50 (1979), 11–26.

[3] FELLER, W., *An Introduction to Probability Theory and its Applications*, vol. 2. Wiley, New York, 1966.

[4] FORD, L. R., *Automorphic Functions*. McGraw-Hill, New York, 1929.

[5] HEJHAL, D., *The Selberg Trace Formula for* PSL(2, **R**). Springer Lecture Notes in Mathematics, 548 (1976).

[6] HUBER, H., Über eine neue Klasse automorpher Functionen und eine Gitterpunktproblem in der hyperbolischen Ebene. *Comment. Math. Helv.*, 30 (1956), 20–62.

[7] HUTCHINSON, J., Fractals and self similarity. *Indiana Univ. Math. J.*, 30 (1981), 713–747.

[8] KATO, T., *Perturbation Theory for Linear Operators*. Springer-Verlag, New York, 1980.

[9] Kolmogorov, A. N. & Tihomirov, V. M., Epsilon-entropy and epsilon-capacity of sets in functional spaces. *Uspekhi Mat. Nauk,* 14 (1959), 3–86.
[10] Krein, M. G., Integral equations on the half-line with a difference kernel. *Uspekhi Mat. Nauk,* 13 (1958), 3–120.
[11] Lalley, S., Distribution of periodic orbits of symbolic and Axiom A flows. *Adv. in Appl. Math.,* 8 (1987), 154–193.
[12] — Regenerative representation for one-dimensional Gibbs states. *Ann. Probab.,* 14 (1986), 1262–1271.
[13] — Packing and covering functions of some self-similar fractals. Unpublished manuscript (1987).
[14] Lax, P. & Phillips, R., The asymptotic distribution of lattice points in euclidean and noneuclidean spaces. *J. Funct. Anal.,* 46 (1982), 280–350.
[15] Lehner, J., *Discontinuous Groups and Automorphic Functions.* Amer. Math. Soc., Providence, 1964.
[16] Mandelbrot, B., *The Fractal Geometry of Nature.* Freeman, New York, 1983.
[17] Margulis, G., Applications of ergodic theory to the investigation of manifolds of negative curvature. *Funktsional. Anal. i Prilozhen,* 3 (1969), 89–90.
[18] Nielsen, J., Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen. *Acta Math.,* 50 (1927), 189–358.
[19] Parry, W., Bowen's equidistribution theorem and the Dirichlet density theorem. *Ergodic Theory Dynamical Systems,* 4 (1984), 117–134.
[20] Parry, W. & Pollicott, M., An analogue of the prime number theory for closed orbits of Axiom A flows. *Ann. of Math.,* 118 (1983), 573–591.
[21] Patterson, S. J., A lattice point problem in hyperbolic space. *Mathematika* 22 (1975), 81–88.
[22] — The limit set of a Fuchsian group. *Acta Math.,* 136 (1976), 241–273.
[23] Pollicott, M., A complex Ruelle-Perron-Frobenius theorem and two counterexamples. *Ergodic Theory Dynamical Systems,* 4 (1984), 135–146.
[24] Rudolph, D., Ergodic behavior of Sullivan's geometric measure on a geometrically finite hyperbolic manifold. *Ergodic Theory Dynamical Systems,* 2 (1982), 491–512.
[25] Ruelle, D., Statistical mechanics of a one-dimensional lattice gas. *Comm. Math. Phys.,* 9 (1968), 267–278.
[26] — *Thermodynamic Formalism.* Addison-Wesley, Reading, 1978.
[27] Series, C., The infinite word problem and limit sets in fuchsian groups. *Ergodic Theory Dynamical Systems,* 1 (1981), 337–360.
[28] — Symbolic dynamics for geodesic flows. *Acta Math.,* 146 (1981), 103–128.
[29] — On coding geodesics with continued fractions. *Enseign. Math.,* 29 (1981), 67–76.
[30] Sullivan, D., On the ergodic theory at infinity of an arbitrary discrete group of hyperbolic motions. *Proc. Stony Brook Conf. Kleinian Groups and Riemann Surfaces,* Princeton University Press (1978).
[31] — The density at infinity of a discrete group of hyperbolic motions. *Inst. Hautes Études Sci. Publ. Math.,* 50 (1979), 419–450.
[32] — Discrete conformal groups and measurable dynamics. *Proc. Sympos. Pure Math.,* 39 (1983), 169–185.
[33] — Entropy, Hausdorff measures new and old, and limit sets of geometrically finite Kleinian groups. *Acta Math.,* 153 (1984), 259–278.

Received November 2, 1987