WILLIAM PARRY
MARK POLLICOTT

Zeta functions and the periodic orbit structure of hyperbolic dynamics

Astérisque, tome 187-188 (1990)

<http://www.numdam.org/item?id=AST_1990__187-188__1_0>

187-188 ASTÉRISQUE

1990

ZETA FUNCTIONS AND THE PERIODIC ORBIT STRUCTURE OF HYPERBOLIC DYNAMICS

William PARRY and Mark POLLICOTT

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

To Benita

and to

Luísa

CONTENTS

page

INTRODUCTION 3

CHAPTER 1 **Subshifts of Finite Type and Function Spaces** 9

Definitions. Hölder continuity. Cohomology of functions. Functions depending on the 'future'.

CHAPTER 2 **The Ruelle Operator** 19

The basic inequality. Ruelle Perron Frobenius Theorem. Strong mixing and exactness. Exponential decay of correlations.

CHAPTER 3 **Entropy, Gibbs Measures and Pressure** 33

Entropy and information. Gibbs and equilibrium states. Variational principle and pressure. Periodic points and cohomology.

CHAPTER 4 **The Complex Ruelle Operator** 49

L^2 versus Hölder for 'maximal' eigenfunctions. Spectral radius of the complex Ruelle operator. Perturbations. Analyticity of pressure. Derivatives of pressure. Variance and the central limit theorem.

CHAPTER 5 **Periodic Points and Zeta Functions** 73

Pressure and the radius of convergence of the zeta function. Cohomology again. The general zeta function and extensions. The main example.

CHAPTER 6 **Prime Orbit Theorems for Suspended Flows** 89

Suspensions of shifts. Associated pressure, equilibrium states and variational principle. L^2 versus Hölder for eigenfunctions. Weak mixing condition. Zeta functions of a suspension and extensions. The shift again and the prime orbit theorem.

CHAPTER 7 **Equidistribution Theorems for Suspended Flows** ... 113

The zeta function of two variables. Proof of Bowen's theorem and its 'weighted' generalisation. The case of zero pressure.

page

CHAPTER 8 **Galois Extensions and Chebotarev Theorems** 123

Hölder continuity for unitary (cocycles) functions. Compact group extensions of suspended flows. Frobenius class. L-functions. Cohomology of unitary functions. A $\mathbb{C}^d$ Ruelle operator and its eigenvalues. Chebotarev theorem. Finite extensions.

CHAPTER 9 **Hyperbolic Flows** 139

Axiom A. Basic sets. Bowen's modelling procedure. Zeta functions. Transference of results to hyperbolic flows. Geodesic flows. Frame bundle flows.

CHAPTER 10 **Further Extension Results** 153

Essential spectrum. The optimal meromorphic extension of the zeta function of a suspension. Examples with an essential singularity. Another counter example.

CHAPTER 11 **Attractors and Synchronisation** 175

Expansion coefficient. Sinai-Ruelle-Bowen measure. Velocity change. Synchronisation. Distribution results related to expansion. The S.R.B. theorem.

CHAPTER 12 **Chebotarev Theorems for $\mathbb{Z}^d$ Extensions** 189

L-functions again. The singularity at $s = 1$. Delange's Tauberian theorem. A Chebotarev theorem. Application to geodesic flows - homology classes. Rees' theorem. Remarks on Poincaré series. The Patterson measure and Sullivan's theorem.

APPENDIX I **The Wiener-Ikehara Tauberian Theorem** 207

APPENDIX II **Unitary Cocycles** 215

APPENDIX III **Hyperbolic Dynamics, Markov Partitions and Zeta Functions** 221

APPENDIX IV **Geodesic Flows** 247

APPENDIX V **Perturbation Theory for Linear Operators** 255

INTRODUCTION

Axiom A diffeomorphisms and flows, introduced by Smale, are generalisations of Anosov systems (extensively studied by the Russian school in the 1960s) which in turn are based on the prototypical hyperbolic toral automorphisms and geodesic flows on surfaces of constant negative curvature. A standard method for understanding these dynamical systems, introduced, at various levels of generality, by Adler, Weiss, Ratner and especially Sinai and Bowen, is to model them, via the introduction of Markov partitions, by shifts of finite type and their suspensions.

In this work we adhere to this procedure and in so doing the initial chapters develop some of the basic material of Bowen's and Ruelle's books [16], [82], although even here many of our proofs are different.

The main substance of our effort, however, takes up where Bowen and Ruelle left off, in that we are primarily interested in problems associated with periodic orbits. These problems are intimately related to the analytic properties of certain zeta functions which may be understood with the help of the Ruelle (Perron-Frobenius) operator. Our point of departure from previous work is, perhaps, our work on the complex Ruelle operator which enables us (at least) to demonstrate the extendibility of a zeta functions up to a certain critical line in the complex plane or even (at best) to obtain Haydn's optimal extension. Crucial to this understanding is an elucidation of the relationship between the spectra of Ruelle operators and the poles of zeta functions.

This volume centres around three theorems which describe, in appropriate settings, the distribution in "space, time and symmetry" of closed orbits for hyperbolic systems. Each of these results is derived by methods inspired by analytic number theory and involves the analysis of a general zeta function.

Temporal Distribution

A zeta function for the closed geodesics of surfaces of constant negative curvature was first put forward by Selberg in 1956 in his work on the trace formula. Huber made implicit use of this work when he established in 1961 an asymptotic formula for the number of closed geodesics and a more general result (for the variable curvature case) was announced by Margulis in 1969. The result we present, proved in 1983, establishes a first order asymptotic for closed orbits of general hyperbolic systems. Our proof, which is entirely analogous to Wiener's proof of the prime number theorem, relies on analyticity properties of the zeta function first defined and partially analysed by Ruelle. This zeta function is a reduced version of the Smale zeta function for flows and is the natural analogue of the Artin-Mazur zeta function for diffeomorphisms, whereas Smale's was inspired by Selberg's.

To be specific the main result on temporal distribution is the following: If φ_t is a topologically weak-mixing hyperbolic flow then the number of closed orbits of least period no more than x is asymptotic to e^{hx}/hx (as $x \to \infty$) where h is the topological entropy of the flow.

Spatial Distribution

In 1972 Bowen proved that closed orbits of an Axiom A flow are uniformly distributed, in the non-wandering set, with respect to a certain canonical measure - the measure of maximum entropy. This result is reproved here, again with the use of a Ruelle zeta function, together with a more general 'weighted' spatial distribution result.

Symmetrical Distribution

Here we take up the work of Sarnak (in his thesis) and Sunada to obtain an analogue of the Chebotarev theorem in number theory. The number theoretical result concerns the distribution of primes according to the way they split in a finite extension field. Our result concerns the distribution of closed orbits according to the way they lift in a Galois extension. Our work generalises the Sarnak-Sunada theorem from geodesic flows to Axiom A flows.

The early chapters include basic material on: shifts of finite type, Hölder continuous functions, the Ruelle-Perron-Frobenius theorem, the Lanford-Ruelle variational principle, pressure, equilibrium states, the central limit theorem, periodic orbits. However, one also finds the less familiar results on the spectral properties of the complex Ruelle operator and an exposition of the analyticity properties of a very general zeta function.

Once we have laid this groundwork we are in a position to prove the temporal (prime orbit), spatial and symmetrical distribution theorems. The final chapters relate results which we have proved for suspended flows (over a shift of finite type) to corresponding ones for hyperbolic flows on manifolds. We also take up a number of miscellaneous themes among which are an optimal meromorphic extension result for the zeta function (due to Haydn), the description of the Sinai-Ruelle-Bowen measure (the 'physical' measure) and a generalisation (due to Adachi and Sunada) of Chapter 8 to $\mathbb{Z}^d$ Galois extensions and its significance for homology.

For the convenience of the reader we conclude with five appendices. The first presents a proof of Ikehara-Wiener Tauberian theorem which enables one to infer asymptotic results from properties of the zeta function. The second concerns a result on unitary cocycles needed for the chapter on Galois extensions. The third is an account of Bowen's theory of Markov partitions including the Bowen-Manning counting lemma and the related correspondence between the zeta function of an Axiom A flow and the zeta functions of associated suspension flows. Appendix IV presents material on geodesic flows and the coding of geodesics and the final appendix gives a brief account of the perturbation theory of linear operators needed for our analysis of Ruelle operators.

Our main aim is to present a reasonably unified account, between one cover, of some of our joint and separate work since 1983 and, of course, to place it in its proper context. Each of us has presented significant portions of this work to graduate classes at the University of Warwick and, in the case of the first named author, at the University of Maryland whereas the second named author presented related material at the California Institute of Technology. We wish to acknowledge

our gratitude to the participants in these courses and seminars. Particular thanks are due to Jawad Al-Khal, Danrun Huang and Marianne James who gave some of the lectures at Maryland and made extensive corrections to preliminary notes for this work. At a later stage our notes benefitted from further corrections due to Anthony Manning, Caroline Series and Richard Sharp for which we extend our gratitude.

David Ruelle very kindly gave us permission to use his rewriting of Haydn's proof of the main theorem in Chapter 10. Almost all of this chapter (with the exception of the example) is a verbatim copy of his notes. Our thanks are also due to the U.K.-Portugal British Council "Treaty of Windsor" for financially supporting the latter part of our joint work.

Above all we thank Alice Gutkind for her typing and for her patience and good humour.

CHAPTER 1

SUBSHIFTS OF FINITE TYPE AND FUNCTION SPACES

We begin by introducing some of the basic objects that we shall need to study. The full shift on k-symbols $(k \geq 2)$ consists in the totality of all doubly infinite sequences of k-symbols together with the shift map (usually denoted σ) which moves each sequence one step to the left. The space of sequences has a natural product topology, and can be viewed as a topological version of an independent (Bernoulli) process. If we specify in advance that a finite number of words (i.e. finite strings of consecutive symbols) shall not be allowed then we obtain a σ-invariant sub-process known as a shift of finite type. It is not difficult to see that if we interpret certain words as new symbols there is no loss in generality if we consider prohibited words of length two. (We need only replace words of a given length by new symbols, cf. [63].)

We shall now be more precise. Let A be a $k \times k$ matrix of zeros and ones $(k \geq 2)$ where the (i,j)th entry is zero precisely when it is a prohibited word of length 2. We define

$$X = X_A = \left\{ x = (x_n)_{n=-\infty}^{\infty} : x_n \in \{1,\ldots,k\}, n \in \mathbb{Z}, A(x_n, x_{n+1}) = 1 \right\}.$$

If $\{1,\ldots,k\}$ is given the discrete topology then X_A is compact and zero-dimensional with the corresponding Tychonov product topology. The *shift*

$\sigma = \sigma_A$ is defined by $\sigma(x) = y$, where $y_n = x_{n+1}$ i.e. all sequences are shifted one place to the left. The pair (X,σ) is called a *shift of finite type* (or *topological Markov chain*).

We shall *always* assume that A is irreducible i.e. for each pair (i,j), $1 \leq i,j \leq k$, there exists $n \geq 1$ such that $A^n(i,j) > 0$, where A^n is an n-fold product of A with itself. Under this condition we define the *period* d of A to be the highest common factor of $\{n : A^n(i,i) > 0, 1 \leq i \leq k\}$. When $d = 1$, A is called *aperiodic.*

There is a unique partition of $\{1,...,k\}$ into sets $S_1,...,S_d$ such that $A^d(i,j) > 0$ only if i,j belong to the same set S_ℓ and A^d is aperiodic when restricted to the index set $S_\ell \times S_\ell$ for each $\ell = 1,...,d$. Moreover, the indexing $S_1,...,S_d$ can be arranged so that if $A(i,j) = 1$ then $i \in S_\ell$, $j \in S_{\ell'}$ where $\ell' = \ell+1$ (mod d). These results are fairly standard and a fuller account may be found in [86].

These properties of the matrix A translate back to the associated *shift of finite type.* It is easy to deduce from these facts that $X = X_A$ can be partitioned into closed-open sets $X = X_1 \cup ... \cup X_d$ so that $\sigma(X_\ell) = X_{\ell'}$ ($\ell' = \ell+1$ mod d) and $\sigma^d \mid X_\ell$ corresponds to an aperiodic matrix. This observation frequently allows us to simplify proofs by replacing the irreducibility hypothesis by the stronger aperiodicity hypothesis and then deducing a more general result, bearing the above comments in mind.

To every (two-sided) shift of finite type we can associate a (*one-sided*) *shift of finite type* $\left(X_A^+, \sigma_A^+\right)$:

$$X^+ = X_A^+ = \left\{x = (x_n)_{n=0}^{\infty}: A(x_n, x_{n+1}) = 1, n \geq 0\right\}$$

and $\sigma^+ x = \sigma_A^+ x = y, y_n = x_{n+1}, n \geq 0$ i.e. all sequences are shifted one place to the left, with the first term being deleted. As before, X^+ is a compact zero-dimensional space with the Tychonov product topology.

An elementary, but important, difference is that whereas the (two-sided) shift $\sigma : X \to X$ is a homeomorphism, the (one-sided) shift $\sigma^+ : X^+ \to X^+$ is not invertible (but merely a local homeomorphism with $\mathrm{Card}(\sigma^+)^{-1}(x) \leq k$). There is a natural continuous surjection $\pi : X \to X^+$ with $\pi(x) = y, y_n = x_n, n \geq 0$ i.e. one deletes the terms $x_n, n < 0$. This surjection clearly satisfies the identity $\pi\sigma = \sigma^+ \pi$.

For a point $x = (x_n)_{n=-\infty}^{\infty} \in X$ we describe $(x_n)_{n=-\infty}^{0}$ as the 'past', x_0 as the present, and $(x_n)_{n=0}^{\infty}$ as the future. To simplify our notation as far as possible we shall write σ for both σ, σ^+. It should always be apparent from the context whether we are referring to a one-sided or a two-sided shift.

Now that we have described the shifts of finite type we can move on to consider function spaces for X, X^+. For future use it transpires that the most interesting family of functions to consider are those that are Hölder continuous.

As before, we begin with X. Given $0 < \theta < 1$ we define a metric on X by $d_\theta(x,y) = \theta^N$, where N is the largest non-negative integer such that $x_i = y_i$, $|i| < N$.

For a continuous function $f : X \to \mathbb{C}$ and $n \geq 0$ we define $\mathrm{var}_n f =$ $= \sup \{|f(x) - f(y)| : x_i = y_i, |i| < n\}$. It is easy to see that $|f(x) - f(y)| \leq Cd_\theta(x,y)$ if and only if $\mathrm{var}_n f \leq C\theta^n$, $n = 0,1,\dots$.

Let $F_\theta = F_\theta(X) = \{f\colon f \text{ continuous}, \mathrm{var}_n f \leq C\theta^n, n = 0,1,\dots, \text{for some } C > 0\}$ then we see that $F_\theta(X)$ is the space of Lipschitz functions with respect to the metric d_θ. For $f \in F_\theta(X)$ let $|f|_\infty = \sup \{|f(x)| : x \in X\}$ and $|f|_\theta = \sup \left\{\dfrac{\mathrm{var}_n f}{\theta^n} : n \geq 0\right\}$. Together these define a norm on F_θ by $\|f\|_\theta = |f|_\infty + |f|_\theta$. (Notice that $|f|_\theta$ is merely the least Lipschitz constant.)

The situation for X^+ is very similar. Given $0 < \theta < 1$ we can define a metric d_θ^+ on X^+ by $d_\theta^+(x,y) = \theta^N$ where N is the largest integer such that $x_i = y_i$, $0 \le i < N$. For a continuous function $f : X^+ \to \mathbb{C}$ and $n \ge 0$ we define $\mathrm{var}_n f =$ $\sup\{|f(x) - f(y)| : x_i = y_i, 0 \le i < n\}$, $|f|_\theta = \sup\left\{\frac{\mathrm{var}_n f}{\theta^n} \mid n \ge 0\right\}$, and $|f|_\infty =$ $\sup\{|f(x)| : x \in X^+\}$. We let

$$F_\theta^+ = F_\theta^+(X^+) = \{f : f \text{ continuous}, \mathrm{var}_n f \le C\theta^n, \ n = 0,1,\ldots, \text{ for some } C > 0\},$$

and again we define a norm on F_θ^+ by $\|f\|_\theta = |f|_\infty + |f|_\theta$.

PROPOSITION 1.1. *The spaces* $(F_\theta, \|\ \|_\theta)$, $(F_\theta^+, \|\ \|_\theta)$ *are Banach spaces. Furthermore, if* $f, g \in F_\theta$ (*or* F_θ^+) *then* $\|fg\|_\theta \le \|f\|_\theta |g|_\infty + |f|_\infty \|g\|_\theta$ *and if* f *is nowhere zero then* $\|1/f\|_\theta \le |1/f^2|_\infty \cdot \|f\|_\theta$.

The proof of this proposition is straight forward. (The proof of completeness is simple since $\{f \in F_\theta : \|f\|_\theta \le C\}$ is $|\ |_\infty$-compact by Ascoli's theorem.)

Two functions $f, g \in F_\theta(X)$ are said to be *cohomologous* $(f \sim g)$ if there exists a continuous function h such that $f = g + h \circ \sigma - h$. Clearly this is an equivalence relation on $F_\theta(X)$. A function which is cohomologous to the zero function is called a *coboundary*. In fact, in the above definition of 'cohomologous' we can always choose $h \in F_\theta(X)$. (Cf. [16].)

PROPOSITION 1.2. *If* $f \in F_\theta(X)$ *then there exist* $g, h \in F_{\theta^{\frac{1}{2}}}(X)$ *such that* $f = g + h - h \circ \sigma$ *and* $g(x) = g(y)$ *whenever* $x_i = y_i$ *for all* $i \geq 0$ (*i.e.* g *depends only on 'future' co-ordinates).*

PROOF. For each $1 \leq j \leq k$ choose an allowable sequence from the 'past' $\left(i_n^j\right)_{n=-\infty}^0$ such that $i_0^j = j$. To each $x \in X$ we can associate $\varphi(x) = x' \in X$ with

$$(x')_n = \begin{cases} x_n, & n \geq 0 \\ i_n^j, & n \leq 0 \text{ and } x_0 = j. \end{cases}$$

Thus φ replaces the 'past' of x by the sequence $\left(i_n^j\right)_{n=-\infty}^0$ where $j = x_0$.

Define $h(x) = \sum_{n=0}^{\infty} (f(\sigma^n x) - f(\sigma^n \varphi x))$. (This series clearly converges since $|f(\sigma^n x) - f(\sigma^n \varphi x)| \leq \mathrm{var}_n f \leq \|f\|_\theta \theta^n, n \geq 0$.) We note that

$$h(x) - h(\sigma x) = \sum_{n=0}^{\infty} (f(\sigma^n x) - f(\sigma^n \varphi x)) - \sum_{n=0}^{\infty} (f(\sigma^{n+1} x) - f(\sigma^n \varphi \sigma x))$$

$$= f(x) - \left[f(\varphi x) + \sum_{n=0}^{\infty} (f(\sigma^{n+1} \varphi x) - f(\sigma^n \varphi \sigma x))\right].$$

This can be rewritten as $h(x) - h(\sigma x) = f(x) - g(x)$, where g is defined by the expression in square brackets. Evidently g depends only on future co-ordinates, and all that remains is to show that h, and therefore g, belongs to $F_{\theta^{\frac{1}{2}}}(X)$. It suffices to show that $\mathrm{var}_{2N} h \leq K\theta^N$, $N \geq 0$, for some constant $K > 0$ for then $\mathrm{var}_{2N+1} h \leq (K/\theta^{\frac{1}{2}})\,(\theta^{\frac{1}{2}})^{2N+1}$.

Let $x, y \in X$ where $x_i = y_i$ for $|i| \leq 2N$ then

$$|f(\sigma^n x) - f(\sigma^n y)|\,,\, |f(\sigma^n \varphi x) - f(\sigma^n \varphi y)| \leq |f|_\theta\, \theta^{2N-n},\ 0 \leq n \leq N.$$

For all $n \geq 0$ we have

$$|f(\sigma^n x) - f(\sigma^n \varphi x)|\,,\, |f(\sigma^n y) - f(\sigma^n \varphi y)| \leq |f|_\theta\, \theta^n .$$

Hence $|h(x) - h(y)| \leq 2|f|_\theta \sum_{n=0}^{N} \theta^{2N-n} + 2|f|_\theta \sum_{n=N+1}^{\infty} \theta^n$

$$= 2|f|_\theta\, \theta^{2N} \left(\frac{\theta^{-N-1} - 1}{\theta^{-1} - 1} \right) + 2|f|_\theta \frac{\theta^{N+1}}{1-\theta} \leq 4|f|_\theta \frac{\theta^N}{1-\theta} ,$$

which shows that $h \in F_{\theta^{\frac{1}{2}}}(X)$, completing the proof.

The map $W : f \mapsto g$ in the above proposition is a linear and continuous map from F_θ to $F_{\theta^{\frac{1}{2}}}$. Furthermore, g can clearly be identified with an element of $F^+_{\theta^{\frac{1}{2}}}$. When $f \in F_\theta$ already depends on future co-ordinates then $W(f) = f$. We can express this as

$$F_\theta(X^+) \hookrightarrow F_\theta(X) \underset{W}{\rightarrow} F_{\theta^{\frac{1}{2}}}(X^+).$$

We have a certain amount of freedom in our choice of $0 < \theta < 1$. Clearly if $f : X \to \mathbb{C}$ is α-Hölder continuous for $d_\theta (0 < \alpha < 1)$ then it is Lipschitz with respect to d_{θ^α} (i.e. replacing θ by θ^α).

Generally, we observe that if $0 < \theta < \theta' < 1$ then $F_{\theta'}(X) \supseteq F_\theta(X)$ (and similarly for F^+_θ, $F^+_{\theta'}$). This gives us a 'filtration' of the spaces of all Hölder continuous functions $F = \bigcup_{0<\theta<1} F_\theta(X)$ (or $F^+ = \bigcup_{0<\theta<1} F_\theta(X^+)$).

Finally, we want to consider a class of functions that lies in all of the F_θ, $0 < \theta < 1$ (or $F^+_\theta, 0 < \theta < 1$). Let

$$F^+_m = \{f : X^+ \to \mathbb{C} : f(x) = f(y) \text{ if } x_n = y_n \text{ for } 0 \le n < m\}$$

for $m \ge 1$ i.e. F^+_m consists of locally constant functions depending on the terms $x_0, \dots, x_{m-1}$. Clearly, $F^+_1 \subseteq F^+_2 \subseteq \dots$ and $\bigcup_{m=1}^{\infty} F^+_m \subseteq \bigcap_{0<\theta<1} F^+_\theta$.

Assume $f \in F_\theta^+$, for some $0 < \theta < 1$, then clearly we can choose $f_m \in F_m^+$ with $|f-f_m|_\infty \leq |f|_\theta . \theta^m$, $m \geq 0$. (In particular, for each admissable word $x_0,\ldots,x_{m-1}$ we can choose $z \in X$ with $z_i = x_i$, $0 \leq i \leq m-1$ and define $f(w) = f(z)$ whenever $w_i = x_i$, $0 \leq i \leq m-1$.)

PROPOSITION 1.3. *For any* $0 < \theta < \theta' < 1$ *we have that* $|f-f_m|_{\theta'} \leq |f|_\theta \left(\frac{\theta}{\theta'}\right)^m$, $m \geq 0$.

PROOF. We want to show $\mathrm{var}_k (f-f_m) \leq |f|_\theta \, \theta^m(\theta')^{k-m}$ for $k \geq 0$.

For the case $0 \leq k \leq m$ we have

$$\mathrm{var}_k (f-f_m) \leq |f-f_m|_\infty \leq |f|_\theta . \theta^m \leq |f|_\theta . \theta^m(\theta')^{k-m}$$

since $(\theta')^{k-m} \geq 1$.

For the case $m < k < +\infty$ we have

$$\mathrm{var}_k (f-f_m) = \mathrm{var}_k f \leq |f|_\theta . \theta^k \leq |f|_\theta . \theta^m(\theta')^{k-m}$$

and the result follows.

In particular, we have $\|f-f_m\|_{\theta'} \leq 2|f|_\theta \left(\frac{\theta}{\theta'}\right)^m$, $m \geq 0$.

Notes

In a purely mathematical context shifts of finite type were introduced in [61] (as "intrinsic Markov chains"). The term "subshift of finite type" was used in [95] whereas the Russian school preferred "topological Markov chain". However shifts of finite type are closely related to the one-dimensional lattice gases extensively studied in statistical mechanics (cf. Ruelle's book [82]).

Details of the arguments about recoding can be found in the books of Parry-Tuncel [64] and Denker-Grillenberger-Sigmund [28]. The reduction to the case of aperiodic matrices is a standard procedure in matrix theory, and a nice account is given in Seneta's book [86].

The importance of Hölder continuous functions on shift spaces is that they correspond to Hölder (or more narrowly, differentiable) functions arising in the context of flows on manifolds. This will be explained in Appendix III. The Banach space of Hölder continuous functions is described in the books of Bowen [16] and Ruelle [82].

Proposition 1.2 is originally due to Sinai [94], but the proof we give is due to Bowen [16].

Proposition 1.3 is taken from Ruelle's book [82].

CHAPTER 2

THE RUELLE OPERATOR

In the previous chapter we introduced the Banach space of Lipschitz functions on subshifts of finite type. For X and X^+ the shift σ induces an operator $\sigma^* : F_\theta \to F_\theta$ or $\sigma^* : F_\theta^+ \to F_\theta^+$. However, in the case of F_θ^+ we have the possibility of introducing an important operator which is dual to σ^*, in a sense which can be made precise.

Let $f \in F_\theta^+$ and define the *Ruelle operator* $L_f : F_\theta^+ \to F_\theta^+$ (or more generally, $L_f : C(X^+) \to C(X^+)$) by $(L_f w)(x) = \sum_{\sigma y = x} e^{f(y)} w(y)$. It is easy to see that L_f is a bounded linear operator. When f is real and $L_f 1 = 1$ we shall sometimes say that f or L_f is *normalised.* Furthermore we have the following:

PROPOSITION 2.1. (Basic inequality) *Let* $f \in F_\theta^+$ *with* $f = u+iv$. *If* $L_u 1 = 1$ *then*

$$|L_f^n w|_\theta \le C|w|_\infty + \theta^n |w|_\theta, \text{ for all } w \in F_\theta^+, n \ge 0$$

where $C > 0$ *depends only on* f *and* θ.

PROOF. We first show that $|L_f w|_\theta \leq C_0 |w|_\infty + \theta |w|_\theta$, for some $C_0 > 0$. Here we choose $d(x,y) \leq \theta^N$ (then $x_i = y_i$ for $0 \leq i < N$), and we note that

$$|(L_f w)(x) - (L_f w)(y)| \leq \sum_{A(i,x_0)=1} |e^{f(ix)}w(ix) - e^{f(iy)}w(iy)|$$

$$\leq \sum_i |e^{f(ix)} - e^{f(iy)}|.|w(ix)| + \sum_i |e^{f(iy)}|.|w(ix) - w(iy)|$$

(where ix denotes the sequence with $(ix)_0 = i$, $(ix)_{n+1} = x_n$, $n \geq 0$), and the result follows easily.

We proceed by induction: If $|L_f^n w|_\theta \leq C_n |w|_\infty + \theta^n |w|_\theta$ then

$$|L_f^{n+1} w|_\theta = |L_f^n (L_f w)|_\theta \leq C_n |L_f w|_\infty + \theta^n |L_f w|_\theta \leq C_n |w|_\infty + \theta^n [C_0 |w|_\infty + \theta |w|_\theta]$$

$$= (C_n + \theta^n C_0)|w|_\infty + \theta^{n+1}|w|_\theta .$$

Thus we can assume $C_{n+1} = C_n + \theta^n C_0 = \left(\sum_{k=0}^{n} \theta^k\right) C_0 = \left(\frac{1-\theta^{n+1}}{1-\theta}\right) C_0 \leq \frac{C_0}{1-\theta}$ and the result is proved if we take $C = \frac{C_0}{1-\theta}$.

The above inequality is the first of two important ingredients in the proof of the theorem below. The second is the elementary observation that $D_1 = \{w \in F_\theta^+ : \|w\|_\theta \leq 1\}$ is compact in the uniform topology, as a subset of $C(X^+)$.

THEOREM 2.2 (Ruelle-Perron-Frobenius, R.P.F.) *Let* $f \in F_\theta^+$ *be real valued and suppose* A *is aperiodic.*

(i) *There is a simple maximal positive eigenvalue* β *of* $L_f : C(X^+) \to C(X^+)$ *with a corresponding strictly positive eigenfunction* $h \in F_\theta^+$.

(ii) *The remainder of the spectrum of* $L_f : F_\theta^+ \to F_\theta^+$ *(excluding* $\beta > 0$*) is contained in a disc of radius strictly smaller than* β.

(iii) *There is a unique probability measure* μ *such that* $L_f^*\mu = \beta\mu$ (*i.e.* $\int L_f v d\mu = \beta \int v d\mu$ *for all* $v \in C(X^+)$).

(iv) $\frac{1}{\beta^n} L_f^n v \to h \int v d\mu$ *uniformly for all* $v \in C(X^+)$ *where* h *is as above and* $\int h d\mu = 1$.

PROOF. Let

$$\Lambda = \{g \in C(X^+) : 0 \le g \le 1 \text{ and } g(x) \le g(y) \exp\left(\frac{|f|_\theta \theta^n}{1-\theta}\right) \text{ when } x_i = y_i\,, 0 \le i \le n\}\,.$$

It is easy to see that Λ is convex and uniformly closed. When $x,y \in X^+$ with $x_i = y_i$, $0 \le i \le n$ we have

$$\begin{aligned}|g(x) - g(y)| &\le |g(y)|\left(\exp\left(\frac{\theta^n}{1-\theta}|f|_\theta\right) - 1\right)\\ &\le |g|_\infty \frac{\theta^n}{1-\theta}\exp\left(\frac{\theta^n}{1-\theta}|f|_\theta\right)|f|_\theta\,.\end{aligned}$$

This allows us to draw two conclusions. The first is that $\Lambda \subset F_\theta^+$. For the second we observe that Λ is an equicontinuous family and by the Ascoli's theorem it is compact with respect to the uniform norm.

For each $n \ge 1$ we may define $L_n(g) = \dfrac{L_f(g+1/n)}{|L_f(g+1/n)|_\infty}$ for $g \in \Lambda$. Clearly $|L_n g|_\infty = 1$ and for $x,y \in X^+$ with $x_i = y_i\,, 0 \le i \le k$,

$$L_f(g+1/n)(x) \le L_f(g+1/n)(y)\exp\left(\frac{\theta^k}{1-\theta}|f|_\theta\right).$$

In particular, $L_n(g)(x) \le L_n(g)(y) \exp\left(\frac{\theta^k}{1-\theta} |f|_\theta\right)$ and so $L_n : \Lambda \to \Lambda$, $n \ge 1$. Since $\Lambda \subset C(X^+)$ is a convex uniformly compact set we can apply the Schauder-Tychonov fixed point theorem to each $L_n : \Lambda \to \Lambda$, $n \ge 1$, to see that there exists $h_n \in \Lambda$ with $L_f(h_n + 1/n) = \beta_n h_n$, where $\beta_n = |L_f(h_n + 1/n)|_\infty$.

By the compactness of Λ we can choose a limit point $h \in \Lambda$ for $\{h_n\}_{n=1}^\infty$ and by continuity $L_f h = \beta h$ where $\beta = |L_f(h)|_\infty$.

To show β is positive we note that

$$\beta_n h_n(x) = \sum_{\sigma y = x} e^{f(y)} (h_n(y) + 1/n) \ge (\inf h_n + 1/n)\, e^{-|f|_\infty}$$

and so $\beta_n(\inf h_n) \ge (\inf h_n + 1/n)\, e^{-|f|_\infty}$. We conclude that $\beta_n \ge e^{-|f|_\infty}$ and so $\beta \ge e^{-|f|_\infty}$.

To show that h is strictly positive we can assume for a contradiction that $h(x) = 0$ for some $x \in X^+$. Then

$$\sum_{\sigma^n y = x} e^{f^n(y)} h(y) = \beta^n h(x) = 0\,,\ n \ge 1$$

where $f^n(y) = f(y) + f(\sigma y) + \cdots + f(\sigma^{n-1} y)$. In particular, $h(y) = 0$ whenever

$\sigma^n(y) = x$, for some $n \geq 0$. Since A is aperiodic the set of all such y is dense in X^+, from which we conclude that h is identically zero. However, we saw above that $\beta = |L_f h|_\infty > 0$, which gives the required contradiction.

To show that β is simple we may suppose that L_f has a second (real-valued) continuous eigenfunction g corresponding to β and let $t = \inf \left\{ \frac{g(x)}{h(x)} \right\} = \frac{g(y)}{h(y)}$, for some $y \in X^+$. Then $g(y) - th(y) = 0$ and $g(x) - th(x) \geq 0$ for all $x \in X^+$. By repeating the preceding argument we conclude that $g - th \equiv 0$, i.e. g is a scalar multiple of h. This shows β is simple, and concludes the proof of part (i).

With h, β as above we define $g = f - \log h \circ \sigma + \log h - \log \beta$, then $L_g = \beta^{-1} \Delta(h)^{-1} L_f \Delta(h)$, where $\Delta(h)$ is multiplication by h. Moreover, $L_g 1 = 1$ so that L_g is normalised. Since the spectrum of L_g is the spectrum of L_f scaled by $1/\beta$, it suffices to complete the proof under the additional assumption that L_f is normalised. The remaining statements then reduce to:

(ii) The spectrum of $L_f : F_\theta^+ \to F_\theta^+$, other than 1, is contained in a disc with radius strictly less than 1.

(iii) There is a unique probability measure m such that $L_f^* m = m$.

(iv) For each $w \in C(X^+)$, $L_f^n w \to \int w dm$, uniformly.

The operator $L_f^* : C(X^+)^* \to C(X^+)^*$ preserves the convex compact (in the weak* sense) subset of functionals corresponding to (σ-invariant) probability measures. In particular, by the Schauder-Tychonov fixed point theorem we can find such an m with $L_f^* m = m$. We complete the proof of (iii) by showing uniqueness and also prove (iv) at the same time.

It is simple to see that $\left\{L_f^n w\right\}$ is equi-continuous, since for all $n,k \geq 1$ we have $\mathrm{var}_k\left(L_f^n w\right) \leq | L_f^n w |_\theta \, \theta^k \leq C\theta^k |w|_\infty + \theta^{n+k} |w|_\theta$ and therefore some convergent subsequence $\left\{L_f^{k_n} w\right\}$ has a limit w^*, say. Since $\sup w \geq \sup L_f w \geq \sup L_f^2 w \geq \cdots$ we have $\sup L_f^N w^* = \sup w^*$, $N = 1,2,\ldots.$ Let $w^*(x_0) = \sup w^* = L_f^n w^*(x_n)$ so that

$$\left(L_f^N w^*\right)(x_N) = \sum_{\sigma^N y = x_N} e^{f^N(y)} w^*(y) = w^*(x_0)$$

then, since L_f is normalised, this is a convex combination and we conclude $w^*(y) = w^*(x_0)$ when $\sigma^N y = x_N$. Thus w^* is a constant.

Since $L_f^* m = m$ we see that $w^* = \int w dm = \lim_{n\to\infty} \int (L_f^{k_n} w)\, dm$. Because $F_\theta^+ \subset C(X^+)$ is uniformly dense we may assume $w \in C(X^+)$. As we may repeat this argument for any subsequence we see that $\lim_{n\to\infty} L_f^n w = \int w dm$ (in the uniform norm). This completes the proof of parts (ii) and (iii).

To prove (ii) it suffices to show $L_f \mid \mathbb{C}^\perp$ has spectral radius strictly less than 1 where:

$$\mathbb{C}^\perp = \left\{ w \in F_\theta^+ : \int w dm = 0 \right\}.$$

By Proposition 2.1 we have

$$| L_f^{n+k} w |_\theta \leq C | L_f^k w |_\infty + \theta^n | L_f^k w |_\theta \leq C | L_f^k w |_\infty + C\theta^n |w|_\infty + \theta^{n+k} |w|_\theta$$

and $L_f^k w$ converges to zero on the uniformly compact set $\{w \in \mathbb{C}^\perp : \|w\|_\theta \leq 1\}$. So for large n,k we have some $\varepsilon > 0$ with $\| L_f^{n+k} w \|_\theta \leq \varepsilon < 1$ for all $w \in \mathbb{C}^\perp$ with $\|w\|_\theta \leq 1$. The spectral radius of $L_f \mid \mathbb{C}^\perp$ is therefore no larger than $\varepsilon^{1/n+k}$ since it is given by $\inf \left\{ \| L_f^N \|_\theta^{1/N} : N \geq 0 \right\}$. This completes the proof.

REMARK 1. (Perron–Frobenius theorem for matrices). *In the special case where* f *depends on only two coordinates, i.e.* $f(x) = f(x_0,x_1)$, *and we can introduce a* $k \times k$-*matrix* $M(i,j) = A(i,j)e^{f(i,j)}$, $1 \le i,j \le k$, *and then* β *is the maximal positive eigenvalue guaranteed by the Perron–Frobenius theorem. In this case* $h(x) = h(x_0)$ *where* $\sum_i h(i)\ A(i,j)e^{f(i,j)} = \beta h(j)$, *i.e.* $h(i)$ *is the ith entry in the eigenvector for* β.

If we define $g(i,j) = \log h(i) - \log h(j) - \log \beta + f(i,j)$ we see that the matrix corresponding to L_g is

$$P(i,j) = A(i,j)e^{g(i,j)} = \frac{A(i,j)\,h(i)e^{f(i,j)}}{\beta h(j)},$$

which is column stochastic, i.e. $\sum_i A(i,j)e^{g(i,j)} = 1$. The measure m on cylinders is given by $m[i_0,i_1,\ldots,i_n] = P(i_0,i_1)\ldots P(i_{n-1},i_n)\,p(i_n)$, where $Pp = p$ and $\sum_i p(i) = 1$ and we use the notation $[i_0,i_1,\ldots,i_n] = \{x \in X^+ : x_j = i_j, j = 0,\ldots,n\}$.

REMARK 2. *Notice that for* $v,w \in L^2(m)$ *we have* $L_f(v.w \circ \sigma) = (L_f v)w$ *so that* L_f *is a partial inverse to the operator* $\sigma^* : w \to w \circ \sigma$ (*when* L_f *is normalised). In particular, we have:* (i) $L_f\sigma^* =$ *identity;* (ii) $\sigma^* L_f = E_m(\cdot \mid \sigma^{-1}\mathcal{B}^+)$ *where* $\mathcal{B}^+$ *is the Borel* σ-algebra *on* X^+ *and* E_m *denotes the conditional expectation for* $\sigma^{-1}\mathcal{B}^+ \subseteq \mathcal{B}^+$. *One should also note that* L_f *is the* $L^2(X^+,m)$ *adjoint of* σ^*.

The measure m satisfying $L_f^* m = m$ when L_f is normalised is clearly σ-invariant since for $v,w \in C(X^+)$ we have $L_f(v \circ \sigma . w) = vL_f w$ and it follows that $\int v \circ \sigma dm = \int v dm$.

We shall also denote by m the natural extension of m from X^+ to X. This is again a σ-invariant measure. (If $v \in C(X^+) \subset C(X)$ we can define $\int v \circ \sigma^{-k} dm = \int v dm$ and note that $\{v \circ \sigma^{-k} : k \geq 0, v \in C(X^+)\}$ is dense in $C(X)$.) Since $L_f \mid \mathbb{C}^\perp$ has spectral radius strictly less than $\rho < 1$ (where $\mathbb{C}^\perp \subset F_\theta^+$), there is a constant $K > 0$ such that $\| L_f^n w \|_\theta \leq K\rho^n \|w\|_\theta$, $n = 0,1,\ldots$ for all $w \in F_\theta^+$ with $\int w dm = 0$.

By considering translates of $F_\theta^+ \subset C(X)$ it is simple to show the following:

PROPOSITION 2.3. *If* $v,w \in L^2(X,m)$ *with* $\int w dm = 0$ *then* $\int v \circ \sigma^n w dm \to 0$ *as* $n \to +\infty$ *(i.e.* σ *is strong mixing with respect to* m).

PROOF. We can choose $v(k), w(k)$, $k \geq 0$ with $\|v(k) - v\|_2$, $\|w(k) - w\|_2 \to 0$ as $k \to +\infty$, where $v(k), w(k)$ depend only on terms $x_{-k},\ldots,x_0,\ldots,x_k$.

We can then write

$$|\int v \circ \sigma^n . w dm| \leq |\int v(k) \circ \sigma^n . w(k) dm| + |\int [v(k) \circ \sigma^n . w(k) - v \circ \sigma^n . w] \, dm|.$$

However, $|\int v(k) \circ \sigma^n . w(k) dm| = |\int [v(k) \circ \sigma^k] \circ \sigma^n [w(k) \circ \sigma^k] \, dm|$

$$= \left|\int [v(k) \circ \sigma^k]\, L_f^n [w(k) \circ \sigma^k]\, dm\right| \le K\rho^n \, \| w(k) \circ \sigma^k \|_\theta \, |v(k) \circ \sigma^k|_\infty$$

which converges to zero as $n \to \infty$ for fixed k. Moreover

$$\left|\int [v(k) \circ \sigma^n . w(k) - v \circ \sigma^n w]\, dm\right| \le \|v\|_2 \, \|w - w(k)\|_2 + \|w(k)\|_2 \, \|v - v(k)\|_2 \,,$$

which can be made arbitrarily small.

REMARK 3. *A simple modification of this argument also shows that* $\sigma : X^+ \to X^+$ *is exact i.e.* $\bigcap_{n=0}^{\infty} \sigma^{-n}\mathcal{B}^+$ *is the trivial* σ-algebra.

If we assume $v, w \in F_\theta$ then we want to examine the rate of convergence to zero.

PROPOSITION 2.4. *If* $v, w \in F_\theta$ *and* $\int w dm = 0$ *then* $\int v \circ \sigma^n . w dm \to 0$ *exponentially fast.*

PROOF. If $v, w \in F_\theta^+ \subset C(X)$ with $\int w dm = 0$ the proof is direct since

$$| L_f^n w |_\infty \le K\rho^n \|w\|_\theta \text{ and so } \left|\int v \circ \sigma^n . w dm\right| \le K\rho^n \|w\|_\theta . |v|_\infty .$$

More generally we need an approximation argument. Let $v,w \in F_\theta^+$ with $\int w dm = 0$ then

$$|L_f^{n+k} w|_\infty \leq K\rho^n \| L_f^k w \|_\theta \leq K\rho^n (|w|_\infty (1+C) + \theta^k |w|_\theta)$$

and if w depends only on the variables $x_0,\dots,x_k$ then

$$|L_f^{n+k} w|_\infty \leq K\rho^n ((1+C)|w|_\infty + 2|w|_\infty) = K'\rho^n |w|_\infty \quad \text{where } K' = K[(1+C)+2]\,.$$

In this case, with the convention $C(X^+) \subset C(X)$,

$$|\int v \circ \sigma^{n+k} w dm| \leq K'\rho^n |w|_\infty . |v|_\infty$$

i.e. $|\int v \circ \sigma^n . w \circ \sigma^{-k} dm| \leq K'\rho^n |w \circ \sigma^{-k}|_\infty |v|_\infty$, where $w \circ \sigma^{-k}$ depends on the variables $x_{-k},\dots,x_0$.

By uniform approximation we have $|\int v \circ \sigma^n . w dm| \leq K'\rho^n |w|_\infty . |v|_\infty$ whenever v depends on future coordinates, w depends on past coordinates and $v,w \in C(X)$ with $\int w dm = 0$.

Returning to the general case, assume $v,w \in F_\theta(X)$ with $\int w dm = 0$ and choose v_k, w_k depending on coordinates $x_{-k},\dots,x_0,\dots,x_k$ with $\int w_k dm = 0$ and

$|v-v_k| \leq |v|_\theta\, \theta^k$, $|w-w_k| \leq |w|_\theta\, \theta^k$. (For example, w_k can be defined by averaging w over cylinders $[x_{-k},\dots,x_0,\dots,x_k]_{-k} = \{z \in X : z_i = x_i,\ -k \leq i \leq k\}$ as a conditional expectation for m.) Then

$$\left|\int v_k \sigma^n . w_k dm\right| = \left|\int v_k \circ \sigma^k \circ \sigma^{n-2k} . w_k \sigma^{-k} dm\right| \leq K'\rho^{n-2k} |v_k|_\infty |w_k|_\infty$$

when $n \geq 2k$ since $v_k \circ \sigma^k$ depends on the future and $w_k \circ \sigma^{-k}$ depends on the past.

Hence for $n \geq 2k$,

$$\begin{aligned}
\left|\int v \circ \sigma^n . w dm\right| &\leq \left|\int (v-v_k) \circ \sigma^n . w dm\right| + \left|\int v_k \circ \sigma^n . (w-w_k)\, dm\right| + \left|\int v_k \circ \sigma^n . w_k dm\right| \\
&\leq \theta^k |v|_\theta\, |w|_\infty + \theta^k |w|_\theta\, |v_k|_\infty + K'\rho^{n-2k} |v_k|_\infty\, |w_k|_\infty \\
&\leq \theta^k |v|_\theta\, |w|_\infty + \theta^k |w|_\theta\, [|v|_\infty + |v|_\theta\, \theta^k] + K'\rho^{n-2k} [|v|_\infty + |v|_\theta\, \theta^k].[|w|_\infty + |w|_\theta\, \theta^k] \\
&\leq \theta^k \|v\|_\theta\, \|w\|_\theta + K'\rho^{n-2k} \|v\|_\theta\, \|w\|_\theta .
\end{aligned}$$

Finally, we can take $k = [n/3]$ then $\left|\int v\sigma^n w dm\right| \leq L\left(\rho^{1/3}\right)^n \|v\|_\theta . \|w\|_\theta$ for some constant $L > 0$ and all $n \geq 0$ (where we assume without loss of generality that $\theta < \rho$).

Notes

The Ruelle operator first appeared as the 'transfer operator' in an article by Ruelle on one-dimensional lattice gases [77], as a generalisation of the 'transfer matrix' - but the related Perron-Frobenius operator is a standard construction.

The basic inequality (Proposition 2.1) is proved in Bowen's book [16]. Inequalities of this type were previously studied by Ionescu-Tulcea-Marinescu [43].

Theorem 2.2 illustrates the reason for introducing the Ruelle operator in preference to the induced operator $\sigma^* : F_\theta^+ \to F_\theta^+$, $(\sigma^* f)(x) = f(\sigma x)$, $x \in X_A^+$. This theorem is due to Ruelle, but we have drawn together proofs of its various parts from different sources: Part (i) uses the proof in Pollicott's article [71]; Part (ii) is taken from Ruelle's book [82]; Part (iii) is adapted from Bowen's book [16] and finally Part (iv) is based on Walters' article [100]. Ledrappier introduces the useful trick of 'normalising', which is closely related to Keane's notion of g-measures [49].

The Perron-Frobenius theorem for matrices can be found in Gantmacher's book [35] and the content of our second remark occurs in an article by Ledrappier [55].

Proposition 2.4 was proved in Bowen's book [16].

CHAPTER 3

ENTROPY, GIBBS MEASURES AND PRESSURE

In this chapter we shall introduce some basic notions from ergodic theory and related ideas originating in statistical mechanics.

Let T be a measure preserving transformation defined on a probability space $(Y,\mathcal{A},p)$ i.e. $T^{-1}\mathcal{A} \subseteq \mathcal{A}$ and $p(T^{-1}A) = p(A)$ for $A \in \mathcal{A}$. If γ is a finite measurable partition and $\mathcal{C} \subset \mathcal{A}$ is a sub-σ-algebra we define the *conditional information of γ given $\mathcal{C}$* as

$$I_p(\gamma|\mathcal{C}) = -\sum_{C\in\gamma}\chi_C \log p(C|\mathcal{C})$$

and the *conditional entropy of γ given $\mathcal{C}$* as

$$H_p(\gamma|\mathcal{C}) = \int I_p(\gamma|\mathcal{C})dp = \int -\sum_{C\in\gamma} p(C|\mathcal{C}) \log p(C|\mathcal{C})dp$$

where $p(C|\mathcal{C}) = E_p(\chi_C|\mathcal{C})$ and we use the convention $x \log x$ is zero at $x = 0$.

The *information* and *entropy* of T with respect to γ are defined, respectively, as $I_p(T,\gamma) = I_p(\gamma|T^{-1}\mathcal{C})$ and $h_p(T,\gamma) = H_p(\gamma|T^{-1}\mathcal{C})$

where $\mathcal{C} = \bigvee_{i=0}^{\infty} T^{-1}\gamma$ is the smallest σ-algebra containing $\bigcup_{i=0}^{\infty} T^{-1}\gamma$.

The *entropy* of T is defined as $h_p(T) = \sup_{\gamma} h(T,\gamma)$ where the supremum is taken over all finite measurable partitions γ. A well-known theorem of Kolmogorov and Sinai asserts $h_p(T) = h(T,\gamma)$ when $\mathcal{A}$ is the smallest T invariant σ-algebra containing γ (i.e. if T is invertible and $\mathcal{A} = \bigvee_{i=-\infty}^{\infty} T^{-1}\gamma$, or more generally if $\mathcal{A} = \bigvee_{i=0}^{\infty} T^{-1}\gamma$).

We can also define $I_p(x) = I_p(\gamma | T^{-1}\mathcal{A})$, when $\mathcal{A}$ is the smallest σ-algebra containing $\bigcup_{i=0}^{\infty} T^{-1}\gamma$, and this is then independent of γ. This definition only has real significance when T is *not* invertible. When T is a continuous surjective map of a compact metric space to itself the *topological entropy* of T is defined as $\sup_p h_p(T)$, where the supremum is taken over all T invariant Borel probabilities p

We shall now restrict ourselves to $(Y,\mathcal{A},p) = (X^+,\mathcal{B}^+,\mu)$ and $T = \sigma$, where $\mathcal{B}^+$ is the Borel σ-algebra for X^+. We can let γ consist of one-cylinders, i.e. $\gamma = \{[i]_0 : i = 1,\ldots,k\}$ then from the above we have $I_\mu(x) = I(\gamma \mid \sigma^{-1}\mathcal{B}^+)$ and $h_\mu(\sigma) = H(\gamma \mid \sigma^{-1}\mathcal{B}^+)$. In this context we can give a convenient expression for these quantities.

For almost all $x \in X^+$ (with respect to μ) we have $\mu[x_0,\ldots,x_n] > 0$. We can define a finite probability distribution on $\{1,2,\ldots,k\}$ for each x,n by

$$\mu_n[i \mid \sigma^{-1}x] = \frac{\mu([i] \cap \sigma^{-1}[x_1,\ldots,x_n])}{\mu(\sigma^{-1}[x_1,\ldots,x_n])}$$

$$= \frac{\mu[i,x_1,\ldots,x_n]}{\mu[x_1,\ldots,x_n]} = \mu\,([i] \mid \sigma^{-1}\gamma \vee\ldots\vee \sigma^{-(n+1)}\gamma)(x)$$

(where $\alpha_1 \vee\ldots\vee \alpha_\ell$ represents the smallest σ-algebra containing all α_i, $i = 1,\ldots,\ell$).

We recall the following:

THEOREM 3.1. (Increasing Martingale Theorem)
With the above notation $\mu_n[i \mid \sigma^{-1}x] \to \mu([i] \mid \sigma^{-1}\mathcal{B}^+)(x)$ *a.e., for each* $i = 1,\ldots,k$, *so that* $\mu[i \mid \sigma^{-1}x] = \mu([i] \mid \sigma^{-1}\mathcal{B}^+)(x)$ *is, for almost all* x, *a well-defined probability distribution on* $1,2,\ldots,k$.

As a consequence we have

$$I_\mu(\sigma)(x) = -\sum_{i=1}^{k} \chi_{[i]}(x) \log \mu[i \mid \sigma^{-1}x]$$

$$h_\mu(\sigma) = \int I_\mu(\sigma)(x)d\mu\,.$$

One can also see that for every $g \in C(X^+)$, $\sum_{i=1}^{k} \int g(ix_1 \ldots)\mu[i \mid \sigma^{-1}x]d\mu = \int g d\mu$.

To show this it suffices to consider $g = \chi_{[j_0 \ldots j_\ell]}$ and note that

$$\sum_{i=1}^{k} \int g(ix_1 \ldots)\, \mu[i \mid \sigma^{-1}x]\, d\mu$$

$$= \lim_{n \to +\infty} \sum_{i=1}^{k} \int g(ix_1 \ldots)\, \mu_n\, [i \mid \sigma^{-1}x]\, d\mu$$

$$= \lim_{n \to +\infty} \sum_{i=1}^{k} \int g(ix_1 \ldots) \frac{\mu[ix_1 \ldots x_n]}{\mu[x_1 \ldots x_n]}\, d\mu$$

$$= \mu\, [j_0 \ldots j_\ell] = \int g d\mu\, .$$

Having introduced some of the more basic ideas from entropy theory we want to relate this to the material in the previous section on the Ruelle operator.

A probability measure m on X^+ is called a *Gibbs measure* if there exists $g \in C(X^+)$ such that

$$A \le \frac{m[x_0, \ldots, x_n]}{e^{g^n(x) + nC}} \le B$$

for $n \ge 0$ and fixed constants $A, B > 0$ and $C \in \mathbb{R}$. Here

$$g^n(x) = g(x) + g(\sigma x) + \cdots + g(\sigma^{n-1}x).$$

(We do not necessarily require that m should be σ-invariant.)

PROPOSITION 3.2. *When* $f \in F_\theta^+$ *is real and normalised we have the following inequality:*

$$e^{-|f|_\theta \theta^n} \le \frac{m[x_0 \dots x_n] e^{-f(x)}}{m[x_1 \dots x_n]} \le e^{|f|_\theta \theta^n}$$

where $L_f^* m = m$ *as in Theorem 2.2.*

PROOF. $m[x_1,\dots,x_n] = \int \chi_{[x_1 \dots x_n]}(z)dm$

$$= \int \sum_{\sigma y = z} \chi_{[x_0 \dots x_n]}(y)dm$$

$$= \int \sum_{\sigma y = z} e^{f(y)} \chi_{[x_0 \dots x_n]}(y) e^{-f(y)} dm$$

$$= \int L_f(\chi_{[x_0 \dots x_n]} e^{-f})(z)dm$$

$$= \int_{[x_0 \dots x_n]} e^{-f} dm .$$

But $e^{-|f|_\theta \theta^n} \le e^{f(z)-f(w)} \le e^{|f|_\theta \theta^n}$ whenever $z,w \in [x_0,\dots,x_n]$. Thus,

$$m[x_0,\dots,x_n]\, e^{-|f|_\theta \theta^n} \le m[x_1,\dots,x_n]\, e^{f(x)} \le m[x_0,\dots,x_n]\, e^{|f|_\theta \theta^n}$$

This completes the proof.

COROLLARY 3.2.1. m *is a Gibbs measure for the constant* $C = 0$.

PROOF. The theorem gives us a sequence of inequalities

$$e^{-|f|_\theta \theta} \leq \frac{m[x_0 \dots x_n]}{m[x_1 \dots x_n]} e^{-f(x)} \leq e^{|f|_\theta \theta^n}$$

$$e^{-|f|_\theta \theta^{n-1}} \leq \frac{m[x_1 \dots x_n]}{m[x_2 \dots x_n]} e^{-f(\sigma x)} \leq e^{|f|_\theta \theta^{n-1}}$$

$$\cdot \qquad \cdot \qquad \cdot$$

$$e^{-|f|_\theta} \leq m[x_n] e^{-f(\sigma^n x)} \leq e^{|f|_\theta} .$$

By multiplying together we have:

$$e^{-|f|_\theta/(1-\theta)} \leq \frac{m[x_0 \dots x_n]}{e^{f^{n+1}(x)}} \leq e^{|f|_\theta/(1-\theta)} .$$

Thus m is a Gibbs measure.

COROLLARY 3.2.2. $\dfrac{m[x_0 \dots x_n]}{m[x_1 \dots x_n]} \to e^{f(x)}$, *uniformly, and* $I_m(\mathcal{B}^+ \mid \sigma^{-1}\mathcal{B}^+) = -f(x)$.

PROOF. The first part is clear from the theorem. Let $\mathcal{B}_n$ be the σ-algebra on X^+ formed from cyclinders of length n, then $I_m\left(\mathcal{B}_n^+ \mid \sigma^{-1}\mathcal{B}_n^+\right)(x) = -\log\left(\frac{m[x_0 \ldots x_n]}{m[x_1 \ldots x_n]}\right)$ and by the martingale theorem $I_m(\mathcal{B}_n^+ \mid \sigma^{-1}\mathcal{B}_n^+) \to I_m(\mathcal{B}^+ \mid \sigma^{-1}\mathcal{B}^+)$ a.e. (m). Combining this with the first part gives $I_m(\mathcal{B}^+ \mid \sigma^{-1}\mathcal{B}^+) = -f(x)$.

These results can easily be adjusted to deal with the case where $f \in F_\theta^+$ and where we no longer necessarily assume that L_f is normalised.

By applying the above theorem to $g = f - \log h \circ \sigma + \log h - \log \beta$, where h, β are the positive eigenfunction and eigenvalue guaranteed by Theorem 2.2 we have

$$A' \leq \frac{m[x_0 \ldots x_n]}{e^{f^n(x) - n\log\beta}} \leq B'$$

for all $x \in X^+$. In particular, m is a Gibbs measure with $C = \log \beta^{-1}$.

If we assume that $f \in F_\theta(X)$ then we can prove similar results by replacing f by a function $g \in F_{\theta^{1/2}}^+ \subset F_\theta(X)$ cohomologous to it.

We next want to consider the way in which invariant Gibbs measures are distinguished amongst all σ-invariant probability measures. This will lead us naturally to consider variational principles. To begin this analysis the following Lemma will prove useful.

LEMMA 3.3. *If* $p_1,\ldots,p_k$ *and* $q_1,\ldots,q_k$ *are two probability distributions on* $1,\ldots,k$ *such that* $p_i > 0, i = 1,\ldots,k$ *then*

$$-\sum_{i=1}^{k} q_i \log q_i + \sum_{i=0}^{k} q_i \log p_i \leq 0$$

with equality only when $p_i = q_i$, $i = 1,\ldots,k$.

PROOF. The left handside of the above inequality can be rewritten as

$$\sum_{i=0}^{k} -p_i \left(\frac{q_i}{p_i}\right) \log \left(\frac{q_i}{p_i}\right)$$

and as the function $\varphi(x) = -x \log x$ (with the convention $\varphi(0) = 0$) is strictly concave, it is less than or equal to $\varphi\left(\sum_{i=1}^{k} p_i \left(\frac{q_i}{p_i}\right)\right) = \varphi(1) = 0$, with equality only when q_i/p_i are all equal $i = 1,\ldots,k$. Hence $q_i = p_i$ for all $i = 1,\ldots,k$.

We use the above lemma as an ingredient in the proof of the following proposition, which gives a preliminary version of the characterisation of m we want.

PROPOSITION 3.4. *If* $f \in F_\theta^+$ *is real with* L_f *normalised and* $L_f^* m = m$, *then for any* σ-invariant *probability measure* μ *we have*

$$h_\mu(\sigma) + \int f d\mu \leq 0$$

with equality if and only if $\mu = m$.

PROOF. Given a σ-invariant probability measure μ we can define a probability distribution on $1,\ldots,k$ by $\mu[i \mid \sigma^{-1}x]$, for almost all $x \in X^+$, with respect to μ. When we choose $\mu = m$ we have $m[i \mid \sigma^{-1}x] = e^{f(ix_0x_1\ldots)}$ for all $x \in X^+$.

In view of Lemma 3.3 we have (for almost all x)

$$-\sum_{i=1}^{k} \mu[i \mid \sigma^{-1}x] \log \mu[i \mid \sigma^{-1}x] + \sum_{i=1}^{k} \mu[i \mid \sigma^{-1}x]\, f(ix_0x_1\ldots) \leq 0$$

with equality a.e. (μ) if and only if $\mu[i \mid \sigma^{-1}x] = e^{f(ix_0x_1\ldots)}$.

Integrating with respect to μ gives:

$$h_\mu(\sigma) + \sum_{i=1}^{k} \int \mu[i \mid \sigma^{-1}x]\, f(ix_0x_1\ldots) d\mu = h_\mu(\sigma) + \int f d\mu \leq 0,$$

with equality if and only if $\mu[i \mid \sigma^{-1}x] = e^{f(ix_0x_1\ldots)}$ a.e. (μ). The latter condition implies $\int\sum_i e^{f(ix_0x_1\ldots)}g(ix_0x_1\ldots)d\mu = \int g d\mu$, when $g \in C(X^+)$ i.e. $\int L_f(g)d\mu = \int g d\mu$ or $L_f^*\mu = \mu$. But by Theorem 2.2 (iii) we know that m is the unique σ-invariant probability measure with $L_f^* m = m$. This completes the proof of the proposition.

We can easily extend the above result to two sided shifts using the correspondences discussed earlier in Theorem 2.2 and Proposition 1.2. Furthermore, we can dispense with the normalisation assumption on the associated Ruelle operator. Thus, by Proposition 1.2 we can find for each $f \in F_\theta(x)$ a function $g \in F_{\theta^{1/2}}(X^+)$ with $f = g + u \circ \sigma - u$. By Theorem 2.2 we can then write $g = \log h \circ \sigma - \log h + \log \beta + k$, where

$L_k^* m = m$ for L_k normalised. By the above proposition for any σ-invariant probability measure μ we have:

$$h_\mu(\sigma) + \int k d\mu \leq h_m(\sigma) + \int k dm = 0$$

i.e. $$h_\mu(\sigma) + \int f d\mu \leq h_m(\sigma) + \int f dm$$

with equality if and only if $\mu = m$. We summarise as follows:

THEOREM 3.5. (Variational Principle) *For* $f \in F_\theta(x)$ *(or* $F_\theta(X^+)$*)*

$$h_\mu(\sigma) + \int f d\mu \leq h_m(\sigma) + \int f dm$$

with equality if and only if $\mu = m$ *for a unique* σ-invariant *probability measure* m.

If we denote $P(f) = \sup_{\mu}\{h_\mu(\sigma) + \int f d\mu\} = h_m(\sigma) + \int f dm$ *then* $P(f) = \log \beta$, *where* β *is the maximal eigenvalue for* $L_{f'}$ *where* $f' \sim f$ *with* $f' \in F^+_{\theta^{1/2}}$.

The quantity $P(f) = \sup_{\mu}\{h_\mu(\sigma) + \int f d\mu\}$ is called the *pressure* of f (and can be similarly defined for any $f \in C(X)$).

A σ-invariant probability measure μ satisfying $P(f) = h_\mu(\sigma) + \int f d\mu$ is called an *equilibrium state.* The above theorem tells us that for $f \in F_\theta$ there exists a unique equilibrium state and that the pressure has an equivalent definition as $P(f) = \log \beta$.

There is a general theory of pressure and equilibrium states for continuous functions with respect to a homeomorphism of a compact metric space, due to Walters [101] and Ruelle [78], which we shall not require.

Next we want to describe some of the basic properties of $P : C(X) \to \mathbb{R}$ with $P(f) = \sup\{h_\mu(\sigma) + \int f d\mu\}$. These are easily seen to follow from the definition:

(i) $P : C(X) \to \mathbb{R}$ is monotone increasing, i.e. if $f,g \in C(X)$, $f \le g$ then $P(f) \le P(g)$;

(ii) $P : C(X) \to \mathbb{R}$ is convex, i.e. for

$0 \leq \lambda \leq 1$, $P(\lambda f + (1-\lambda)g) \leq \lambda P(f) + (1-\lambda)P(g)$;

(iii) If $f \sim g + c$, for some constant c, then $P(f) = P(g) + c$;

(iv) $P : C(X) \to \mathbb{R}$ is Lipschitz continuous, i.e. $|P(f) - P(g)| \leq |f-g|_\infty$. For if $|f-g|_\infty = c$ then $g-c \leq f \leq g+c$ so that $P(g) - c \leq P(f) \leq P(g) + c$ by (i) and (iii) above. Thus $|P(f) - P(g)| \leq |f-g|_\infty = c$.

The following result shows that there is a one-one correspondence between elements of $F_\theta(X)$ (modulo coboundaries plus constants) and equilibrium states of F_θ functions.

PROPOSITION 3.6. *If* $f,g \in F_\theta(X)$ *and* $f \sim g + c$, *where* c *is a constant then* f *and* g *have the same equilibrium state. Conversely, if* f *and* g *have the same equilibrium state then* $f \sim g + c$, *where* c *is constant.*

PROOF. For the first part we note that the equilibrium state m of f is defined by $h_m(\sigma) + \int f dm = P(f)$. Therefore $h_m(\sigma) + \int g dm = h_m(\sigma) + \int f dm - c = P(f) - c = P(g)$ i.e. m is the equilibrium state for g.

For the second part, let m be the common equilibrium state of f and g. Since we are only interested in equating f and g up to the addition of a coboundary and a constant we can assume that L_f, L_g are normalised (using Proposition 1.2 and Theorem 2.2). Thus $L_f 1 = 1$, $L_g 1 = 1$ and $L_f^* m = m$, $L_g^* m = m$. In particular, $\int L_f w dm = \int w dm = \int L_g w dm$ for all $w \in C(X^+)$. If we let $w = u.v \circ \sigma$ then $\int v\, L_f u\, dm = \int v\, L_g u\, dm$. Thus $L_f = L_g$ from which it follows $e^{f(x)} = L_f \chi_{[x_0]}(\sigma x) = L_g \chi_{[x_0]}(\sigma x) = e^{g(x)}$ i.e. $f \equiv g$ (when L_f, L_g are normalised).

The above proposition shows that we may recover $f \in F_\theta(X)$ (up to a coboundary and a constant) from its equilibrium state. The next result is in a similar spirit, and essentially says that $f \in F_\theta(X)$ is determined uniquely (up to a coboundary) by the sum of its values around periodic orbits. This result, due to Livsic, will be considered again in Chapter 5.

PROPOSITION 3.7. (Livsic [56]) *Two functions* $f,g \in F_\theta(X)$ *satisfy* $f \sim g$ *if and only if* $f^n(x) = g^n(x)$ *whenever* $\sigma^n x = x$.

PROOF. It suffices to show that if $f^n(x) = 0$ whenever $\sigma^n x = x$ then f is a coboundary. Fix $x_0 \in X$ such that $\left\{\sigma^n x_0\right\}_{n=0}^{\infty}$ is dense in X. We want to define $u \in C(X)$ by $u(\sigma^n x_0) = f^n(x_0)$ on this dense orbit. If $y = \sigma^n x_0$ and $y' = \sigma^{n+m} x_0 = \sigma^m y$ then we can choose a periodic point $\sigma^m x = x$ with $x_i = y_i$, $i = 0,\dots,m-1$, provided y and y' are sufficiently close. Assume that $d(y,y') \le \theta^k$, say, then we have

$$\begin{aligned}|u(y) - u(y')| &= |f(\sigma^{n+m-1}x_0)+\cdots+f(\sigma^n x_0)| \\ &= |f^m(y) - f^m(x)| \\ &\le \sum_{i=0}^{m-1} |f(\sigma^i y) - f(\sigma^i x)|\end{aligned}$$

(where we have used the fact that $f^m(x) = 0$).

Since $d(\sigma^m y, y) = d(y', y) \le \theta^k$ we see that $y_i = y_{i+m}$, for $-k \le i \le k$. In particular, $x_i = y_i$, for $|i| \le m + k$.

Thus $|f(\sigma^i y) - f(\sigma^i x)| \le |f|_\theta\, \theta^{k+m-i}$, for $i = 0,\dots,m-1$ and hence

$$|u(y) - u(y')| \le \sum_{i=0}^{m-1} |f|_\theta\, \theta^{k+i} \le \frac{|f|_\theta}{1-\theta}\, \theta^k .$$

Thus u extends to a continuous function on X. (In fact, we can see that

$u \in F_\theta$.)

If $y = \sigma^n x_0$ then $u(\sigma y) - u(y) = f^{n+1}(x_0) - f^n(x_0) = f(\sigma^n x_0) = f(y)$. This identity extends to all $y \in X$ by continuity.

Notes

Some basic ideas and results from ergodic theory can be found in [62], including a proof of the increasing martingale theorem (Theorem 3.1).

The introduction of Gibbs measures into ergodic theory and dynamical systems began with the important paper of Sinai [94]. Expositions of this theory occur in the books of Bowen [16] and Ruelle [82]. Our treatment is probably closer to that of Bowen, at least in terms of notation.

The variational principle (Theorem 3.5) originated in statistical mechanics. See, for example, the work of Lanford and Ruelle [53]. Alternatives to our proof occur in the books of Bowen [16] and Ruelle [82].

Following work of Ruelle [78], Walters produced the most general version of the variational principle for homeomorphisms of compact metric spaces [101] (and there is a shortened proof due to Misiurewicz [58]).

Basic properties of pressure, Gibbs measures and equilibrium states can be found in the books of Bowen [16], Walters [102] and Ruelle [82]. Ruelle's book also describes a parallel theory in which functions are replaced by 'interactions'.

Proposition 3.6 was originally proved by Ruelle [82] using a different approach involving "strict convexity of pressure".

Livsic's theorem (Proposition 3.7) is taken from [56].

CHAPTER 4

THE COMPLEX RUELLE OPERATOR

So far we have considered the Ruelle operator L_f for f real-valued and developed it as a tool to study pressure, equilibrium states, etc. In this chapter we want to consider L_f with $f \in F_\theta(X^+) = F_\theta(X^+, \mathbb{C})$. We then refer to $L_f : F_\theta(X^+) \to F_\theta(X^+)$ as the *complex Ruelle operator.* We shall assume that σ is aperiodic.

For convenience, we can assume that if $f = u + iv$ then L_u is normalised, i.e. $L_u 1 = 1$. (Here $u,v \in F_\theta(X^+, \mathbb{R})$ are the real and imaginary parts of f.) We let m denote the equilibrium state of u (in particular, $L_u^* m = m$).

In view of the basic inequality (Proposition 2.1), we have

$$\| L_f^n w \|_\theta \leq (C+1)\, |w|_\infty + \theta^n\, |w|_\theta \leq (C+1)\, \|w\|_\theta \, .$$

Thus by the spectral radius theorem we see that $L_f : F_\theta(X^+) \to F_\theta(X^+)$ has spectral radius at most unity.

Our main concern will be the possibility that L_f has an eigenvalue of modulus one. To investigate this phenomenon it is convenient to introduce an operator $Vw = e^{-iv} w \circ \sigma$. This is defined on any of the three spaces $L^2(m)$, $C(X^+)$ or $F_\theta(X^+)$, and is an isometry on the first two. The basic relationship between L_f and V is that $L_f V w = w$, $VL_f w = M_v^{-1} . E_m(w \mid \sigma^{-1} \mathcal{B}^+)) . M_v$ where $M_v w = e^{iv} w$.

PROPOSITION 4.1. *Either* $\bigcap_{n=0}^{\infty} V^n L^2(m)$ *is trivial or it is one-dimensional. Furthermore, the intersection is one-dimensional if and only if* V *has a simple eigenvalue in* $L^2(m)$.

PROOF. Assume that $\bigcap_{n=0}^{\infty} V^n L^2(m)$ is non-trivial, and let $w \neq 0$ lie in this intersection. We can write $w = Vw_1 = \dots = V^n w_n = V^{n+1} w_{n+1}$, $n = 0,1,\dots$, so that

$$w = e^{-iv^n} w_n \circ \sigma^n = e^{-iv^{n+1}} w_{n+1} \circ \sigma^{n+1} .$$

Hence, $w \circ \sigma e^{-iv} = e^{-iv^{n+1}} w_n \circ \sigma^{n+1}$ and $w \circ \sigma e^{-iv}/w = w_n \circ \sigma^{n+1}/w_{n+1} \circ \sigma^{n+1}$ is $\sigma^{-(n+1)} \mathcal{B}^+$-measurable for each $n = 0,1,2,\dots$. (We observe that $w \circ \sigma . e^{-iv}/w$ is well defined since $|w|$ is $\sigma^{-n} \mathcal{B}^+$ measurable for $n \geq 0$. In particular, since σ is an exact endomorphism $|w|$ must be constant and non-zero.)

Since σ is exact we conclude that $w \circ \sigma . e^{-iv}/w$ is constant and $w \circ \sigma . e^{-iv} = \alpha w$ for a non-zero function w. If we have a second solution $w' \circ \sigma . e^{-iv} = \alpha' w'$

then w'/w is an eigenfunction for σ^*. However, $\sigma : X^+ \to X^+$ is mixing so that the only eigenfunctions are the constants, i.e. $w = cw'$ for some constant c. Thus the intersection is one-dimensional, completing the proof of the first statement. In the course of this argument we saw that non-triviality (or equivalently, one-dimensionality) of the intersection implies that V has a simple eigenvalue. This completes the proof.

PROPOSITION 4.2. V *has an* $L^2(m)$ *eigenfunction (or equivalently,* $\bigcap_{n=0}^{\infty} V^n L^2(m)$ *is one-dimensional) if and only if* V *has an* F_θ^+ *eigenfunction.*

PROOF. Let $w \in L^2(m)$ $(|w| = 1)$ be an eigenfunction for V then $w \circ \sigma . e^{-iv} = \alpha w$, say. As we observed in the proof of the previous proposition w is non-zero so we can write $w \circ \sigma^n / w = \alpha^n e^{iv^n}$. Hence $wL_u^n(g/w) = \alpha^n L_f^n g$ for any $g \in F_\theta^+$. Since $|\alpha| = 1$ and $L_f^n g$ is $|\ |_\infty$ equicontinuous (by the basic inequality, Proposition 2.1) we can choose subsequences with α^{n_k} converging to α^*, say, and $L_f^{n_k} g$ converging uniformly to $g^* \in F_\theta^+$.

In view of Theorem 2.2 (iv) we see that $L_u^n(\bar{w}g)$ converges to $\int \bar{w} g dm$. This gives the equation

$$w\int \bar{w} g dm = \alpha^* g^* .$$

Since w is assumed to be non-trivial we may choose g with $\int \bar{w} g dm \neq 0$ and conclude that w is a scalar multiple of g^* a.e. (m). So the eigenfunction of V may be chosen in F_θ^+.

PROPOSITION 4.3. *If* V *has no eigenfunctions (in* $L^2(m)$, *or equivalently,* F_θ^+) *then* $L_f^n g$ *converges uniformly to zero for all* $g \in F_\theta^+$.

PROOF. Again we can use the equicontinuity of $\{L_f^n g\}, g \in F_\theta^+$, to choose a uniformly convergent subsequence $L_f^{n_k} g \to g^*, g^* \in F_\theta^+$.

If $w \in L^2(m)$, then $\int w . L_f^{n_k} g \, dm \to \int w g^* dm$, i.e. $\int V^{n_k} w . g dm \to \int w g^* dm$. Since each $V^{n_k} w$ has the same $L^2(m)$ norm there is a further subsequence (which we again denote $V^{n_k} w$) such that $V^{n_k} w$ converges weakly in $L^2(m)$. As this weak limit lies in $\bigcap_{n=0}^{\infty} V^k L^2(m)$ it must be zero by Proposition 4.1, for we have assumed V has no eigenfunctions. Thus $\int w g^* dm = 0$ for arbitrary $w \in L^2(m)$ and so $g^* \equiv 0$. Thus we can conclude that $L_f^n g \to 0$ uniformly for $g \in F_\theta^+$.

PROPOSITION 4.4. *If* V *has no eigenfunctions (in* $L^2(m)$ *or, equivalently, in* F_θ^+) *then* $L_f^n \to 0$ *in the* $\|\ \|_\theta$*-operator topology, i.e.* L_f *has spectral radius* $\rho(L_f) < 1$.

PROOF. Let $w \in D_1 = \{w : \|w\|_\theta \leq 1\}$ then $|L_f w|_\infty \leq |w|_\infty$ and by the previous proposition $|L_f^n w|_\infty$ converges to zero uniformly for functions in D_1, i.e. for all $\varepsilon > 0$ there exists $N > 0$ such that $|L_f^n w|_\infty \leq \varepsilon$ for all $w \in D_1$ and all $n \geq N$.

Moreover, by the basic inequality (Proposition 2.1) applied twice:

$$\begin{aligned}
|L_f^{n+N} w|_\theta &\leq C|L_f^N w|_\infty + \theta^n |L_f^N w|_\theta \\
&\leq C|L_f^N w|_\infty + \theta^n (C|w|_\infty + \theta^N |w|_\theta) \\
&\leq C\varepsilon + \theta^n (C+\theta^N) \\
&\leq (2C+1)\varepsilon < 1
\end{aligned}$$

(provided $\varepsilon < 1/(2C+1)$) for all $w \in D_1$, when n is large. Thus $\| L_f^{n+N} \|_\theta < 1$, and the proposition is proved.

The final thing we want to do is to relate the eigenvalue condition for V to one for L_f. In particular, V has an $L^2(m)$ or F_θ^+ eigenvalue if and only if L_f has an $L^2(m)$ or F_θ^+ eigenvalue of *modulus 1.*

To see this, assume first that $Vw = \alpha w$ where $w \in F_\theta^+$ (or equivalently $L^2(m)$) and α is necessarily of modulus one. Thus $w \circ \sigma = \alpha e^{iv} w$ and therefore $L_f w = \bar{\alpha} w$. Conversely, assume that $L_f w = \bar{\alpha} w$ ($|\alpha| = 1$) then $L_u |w| \geq |w|$ a.e. (m), where $w \in L^2(m)$. Since integration by m with $L_u^* m = m$ implies $L_u |w| = |w|$ a.e. we conclude that $|w|$ is constant a.e. (m). Because $L_u 1 = 1$ and $L_u(e^{iv} w) = \bar{\alpha} w$ we can use a convexity argument to deduce that $e^{iv(y)} w(y) = \bar{\alpha} w(x)$ for all y with $\sigma y = x$, for almost all x. Thus $Vw = \alpha w$, $w \in L^2(m)$. By Proposition 4.2 we can assume $w \in F_\theta^+$.

This brings us to the main result of this chapter.

THEOREM 4.5. *For* $f = u + iv \in F_\theta^+$ *we have* $\rho(L_f) \leq e^{P(u)}$. *If* L_f *has an eigenvalue of modulus* $e^{P(u)}$ *then it is simple and unique and* $L_f = \alpha M L_u M^{-1}$, *where* M *is a multiplication operator and* $\alpha \in \mathbb{C}$, $|\alpha| = 1$. *Furthermore, the rest of the spectrum is contained in a disc of radius strictly smaller than* $e^{P(u)}$. *If* L_f *has no eigenvalues of modulus* $e^{P(u)}$ *then the spectral radius of* L_f *is strictly less than* $e^{P(u)}$.

PROOF. Using Proposition 1.2 and Theorem 2.2 we can write $u = u' + w\sigma - w + P(u)$, where $L_{u'}$ is normalised. If we let M represent multiplication by e^w then this becomes $L_f = e^{P(u)} M L_g M^{-1}$, where $g = u' + iv$.

The condition that L_f has an eigenvalue of modulus $e^{P(u)}$ is equivalent to L_g having an eigenvalue of unit modulus. If $L_g w = \alpha w$, $w \in F_\theta^+$, $|\alpha| = 1$ then by our previous comments we have $|w| = 1$, say, and $e^{iv} = \alpha w \circ \sigma / w$ and therefore

$$L_g q = L_{u'}(e^{iv} q) = \alpha w L_{u'}(w^{-1}.q)\ , \ \text{for any } q \in F_\theta^+ .$$

The spectral properties of L_g (and hence L_f) follow from those of L_u described in Theorem 2.2 (ii).

The condition that L_f has no eigenvalues of modulus $e^{P(u)}$ is equivalent to L_g having no eigenvalues of unit modulus. The condition $\rho(L_f) < e^{P(u)}$ comes from $\rho(L_g) < 1$ by Proposition 4.4.

REMARK. *In the case where* $f = f(x_0, x_1)$ *depended on only two coordinates and* f *was real-valued, Theorem 2.2 reduced to the familiar Perron-Frobenius theorem for matrices. If we assume in the above theorem that* $f = f(x_0, x_1)$ *we can reduce the statement to Weilandt's theorem for the matrix* $M(i,j) = A(i,j)\, e^{f(i,j)}$, $1 \le i,j \le k$: *Let* N *be the positive matrix with* $N(i,j) = |M(i,j)| \ge 0$ *and let* $\lambda > 0$ *be the maximal positive eigenvalue for* N. *The eigenvalues for* M *all have moduli strictly less than* λ *unless* M *has the form* $M = e^{i\theta} U N U^{-1}$, *where* $0 \le \theta < 2\pi$ *and* $U = \mathrm{diag}\,(e^{i\theta_1}, \dots, e^{i\theta_k})$ *with* $0 \le \theta_1, \dots, \theta_k \le 2\pi$. (An account of this result can be found in Gantmacher [35].)

In the previous chapter (Theorem 3.5) we interpreted the maximal positive eigenvalue for L_f, when $f \in F_\theta$ is real valued, in terms of the pressure $P(f)$. The last theorem gives us a way of extending the definition to certain complex valued functions and studying its regularity. The advantage of defining pressure in terms of Ruelle operators is that we can make use of some standard results from the perturbation theory of linear operators. The following result is particularly useful.

PROPOSITION 4.6. (Perturbation theorem). *Let* $B(V)$ *denote the Banach algebra of bounded linear operators on a Banach space* V. *If* $L_0 \in B(V)$ *has a simple isolated eigenvalue* α_0 *with corresponding eigenvector* v_0 *then for any* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that if* $L \in B(V)$ *with* $\|L-L_0\| < \delta$ *then* L *has a simple isolated eigenvalue* $\alpha(L)$ *and corresponding eigenvector* $v(L)$ with $\alpha(L_0) = \alpha$, $v(L_0) = v_0$ *and*

(i) $L \mapsto \alpha(L), L \mapsto v(L)$ *are analytic for* $\|L-L_0\| < \delta$

(ii) *for* $\|L-L_0\| < \delta$, *we have* $|\alpha(L) - \alpha_0| < \varepsilon$, *and spectrum* $(L) - \{\alpha(L)\} \subseteq \{z : |z-\alpha_0| > \varepsilon\}$.

(For a more detailed discussion of perturbation theory we refer to [8] or [44]. See also Appendix V.)

We can extend the definition of pressure to functions $f \in F_\theta^+$ (for some $0 < \theta < 1$), with the property that $L_f : F_\theta^+ \to F_\theta^+$ has a simple 'maximum' eigenvalue λ such that the rest of the spectrum of L_f is contained in a disc with radius strictly less than $|\lambda|$. For such functions f we extend the definition of pressure by $P(f) = \log \lambda$. (Formally this definition can only be made modulo $2\pi i$ since log is multiple valued, although we shall ask that $P(f)$ be real-valued when f is real-valued.)

Locally $f \mapsto P(f)$ is well-defined. Furthermore, $P(f) = P(g) + c$ whenever $f, g \in F_\theta^+$ and $f \sim g + c + 2\pi i M$ where M is continuous and integer valued and c constant.

PROPOSITION 4.7. *The domain of* P *(denoted* $\mathrm{dom}\,(P) \subseteq F_\theta^+$*) is open and* $f \mapsto P(f)$ *is an analytic map from* $\mathrm{dom}\,(P)$ *into* $\mathbb{C}$.

PROOF. We need only prove analyticity. Since the perturbation theorem states that $L \mapsto \alpha(L)$ is analytic on the open set where it is defined it suffices to show that the map $f \mapsto L_f$, $F_\theta^+ \to B(F_\theta^+)$ is analytic.

Consider the composition of maps, $F_\theta^+ \to F_\theta^+ \to B(F_\theta^+) \to B(F_\theta^+)$ given by $f \mapsto e^f \mapsto M \mapsto L_i \circ M$, where M is the multiplication (by e^f) operator, and

$$(L_i w)(x) = \begin{cases} w(ix) & \text{if } A(i,x_0) = 1 \\ 0 & \text{otherwise}. \end{cases}$$

Each of these maps can be seen to be analytic. Finally, we note that $L_f = \sum_{i=1}^{k} L_i \circ M$ and conclude that $f \mapsto L_f$ is analytic.

We can naturally define an extension of $P : F_\theta \to \mathbb{R}$ using the above extension of $P : F_\theta^+ \to \mathbb{R}$ to $P : \mathrm{dom}(P) \to \mathbb{C}$. Let $W : F_\theta \to F^+_{\theta^{1/2}}$, $f \mapsto Wf$ be the linear map from Proposition 1.2 for which $f = Wf + f' \circ \sigma - f'$ (with $f' \in F_{\theta^{1/2}}$ and $W \equiv I$ on $F_\theta^+ \subset F^+_{\theta^{1/2}}$).

We define $\mathrm{Dom}(P) = W^{-1}\,\mathrm{dom}(P) \subset F_\theta(X,\mathbb{C})$, and $P : \mathrm{Dom}(P) \to \mathbb{C}$, $f \mapsto P(Wf)$. Since the choice of W in Proposition 1.2 is not unique we want to show that this extension of P is independent of the specific choice. This requires showing that if $f = g + w \circ \sigma - w$ with $f,g \in F_\theta^+$ and $w \in F_\theta$ then $w \in F_\theta^+$.

If $x_n = y_n$, $n \geq 0$ then
$$w(\sigma x) - w(x) = w(\sigma y) - w(y)$$
$$w(\sigma^2 x) - w(\sigma x) = w(\sigma^2 y) - w(\sigma y).$$

Thus for all $n \geq 0$, $w(\sigma^n x) - w(\sigma^n y) = w(x) - w(y)$. Since $d(\sigma^n x, \sigma^n y) \to 0$ we have $w(x) = w(y)$, i.e. $w \in F_\theta^+$. We can also see from the definition that when $f \sim g + c + 2\pi i M$ (M integer valued) then $P(f) = P(g) + c$. We summarise as follows:

PROPOSITION 4.8. *The domain of* P ($\mathrm{Dom}\ P \subset F_\theta$) *is open and* $f \mapsto P(f)$ *is an analytic map from* $\mathrm{Dom}\ P$ *into* $\mathbb{C}$ *such that* $f \sim g + c + 2\pi i M$ (M *integer valued) implies that* $P(f) = P(g) + c \pmod{2\pi i}$.

The extended definition of pressure leads quite simply to the following results.

PROPOSITION 4.9. *If* $f \in \mathrm{Dom}\ P$ *then* $P(\mathcal{R}f) \geq \mathcal{R}P(f)$ *with equality if and only if* f *is cohomologous to* f' with $\rho(L_{f'}) = e^{P(\mathcal{R}f)}$.

In the remainder of this chapter we shall be largely concerned with real functions $f, g \in F_\theta^+$ with f normalised and with $\int g\,dm = 0$ where m is the equilibrium state of f. In this situation we shall consider the perturbations of L_f given by L_{f+sg} (s small) and we shall need to be more precise about the corresponding perturbations of the maximum eigenvalue 1 and the associated eigenfunction 1 for the operator L_f.

Since 1 is a simple isolated eigenvalue of L_f, we have a projection valued analytic function $Q(s)$, defined for small complex s, such that

$$L_{f+sg}Q(s) = Q(s)L_{f+sg}$$

and hence

$$L_{f+sg}w(s) = e^{P(f+sg)}w(s) \tag{4.1}$$

where $w(s,x) = w(s) = Q(s)1$. Hence $e^{P(f+sg)}$ is a (maximum) simple isolated eigenvalue for the operator L_{f+sg} if s is small.

Differentiating both sides of (4.1) at $s = 0$ and integrating with respect to m yields

PROPOSITION 4.10. *If* $f,g \in F_\theta$ *are real and if* m *is the equilibrium state of* f *then*

$$P'(0) = \frac{dP(f+sg)}{ds}\Big|_{s=0} = \int g dm .$$

PROOF. It suffices to note that there is no loss in generality in assuming that $f,g \in F_\theta^+$ and $\int gdm = 0$, and the above computation shows that $P'(0) = 0$.

A second differentiation at $s = 0$ yields, after integration,

$$\int g^2 dm + 2\int gw'(0)dm = P''(0)$$

and the same steps applied to

$$L^n_{f+sg}\, w(s) = e^{nP(f+sg)} w(s)$$

leads to

$$\int (g^n)^2 dm + 2\int g^n . w'(0) dm = nP''(0)$$

so that a simple application of the ergodic theorem gives

PROPOSITION 4.11. *If* $f, g \in F_\theta$ *are real and if* $\int g dm = 0$ *where* m *is the equilibrium state of* f *then*

$$P''(0) = \lim_{n\to\infty} \frac{1}{n} \int (g^n)^2 dm\,.$$

PROOF. Modify f,g by the addition of coboundaries, for convenience, and apply the procedure preceding the proposition.

If $f, g \in F_\theta$ are real and m is the equilibrium state of f we define the *variance* of the process $\{g \circ \sigma^n\}$ with respect to f (or m) by

$$\sigma_f^2(g) \equiv P''(0) = \frac{d^2 P(f+tg)}{dt^2}\bigg|_{t=0} = \frac{d^2 P(f+t(g-\int g dm))}{dt^2}\bigg|_{t=0}$$

$$= \lim_{n\to\infty} \frac{1}{n} \int (g^n - n\int g\,dm)^2 dm\,.$$

We are now in a position to prove:

PROPOSITION 4.12. *Let* $f,g \in F_\theta$ *then* $\sigma_f^2(g) \geq 0$ *with equality if and only if* g *is cohomologous to a constant. Hence* $t \to P(f+tg)$ *is convex - and strictly convex if* $\sigma_f^2(g) > 0$.

PROOF. Evidently the last statement follows from the first for if

$$\left.\frac{d^2P(f+tg)}{dt^2}\right|_{t=t_0} = 0 \text{ then } \sigma^2_{f+t_0g}(g) = 0$$

so that g is cohomologous to a constant and then

$$\frac{d^2P(f+tg)}{dt^2} = 0\,,$$

for all $t \in \mathbb{R}$.

It is clear that if g is cohomologous to a constant then $\sigma_f^2(g) = 0$. We have to prove the converse. We assume $\int g\,dm = 0$.

Suppose now that $\sigma_f^2(g) = 0$. By Herglotz's theorem and exponential convergence of correlations we may write

$$\int g \circ \sigma^n g \, dm = \int_K \lambda^n r(\lambda) d\lambda$$

where r is analytic on the circle K. Thus

$$\begin{aligned} 0 &= \lim_{n\to\infty} \frac{1}{n} \int (g^n)^2 dm = \lim_{n\to\infty} \frac{1}{n} \int | 1 + \cdots + \lambda^{n-1} |^2 r(\lambda) d\lambda \\ &= \lim_{n\to\infty} \frac{1}{n} \int \left| \frac{1-\lambda^n}{1-\lambda} \right|^2 r(\lambda) d\lambda = r(1) \end{aligned}$$

by a well known property of the Fejer kernel. Also,

$$r(\lambda) = \int g^2 dm + \sum_{n=1}^{\infty} (\lambda^n + \lambda^{-n}) \int g \circ \sigma^n g dm$$

and differentiation at $\lambda = 1$ gives $r'(1) = 0$.

Since $r(\lambda)$ is analytic at $\lambda = 1$ we therefore have, for $\lambda - 1$ small,

$$r(\lambda) = (\lambda - 1)^2 s(\lambda)$$

with $s(\lambda)$ analytic at $\lambda = 1$. In particular

$$\int \frac{r(\lambda)}{|\lambda-1|^2}\, d\lambda < \infty \text{ and } \frac{r^{1/2}(\lambda)}{\lambda-1} \in L^2(K).$$

This shows that $r^{1/2}(\lambda)$ is a coboundary with respect to the unitary operator given by multiplication by λ. The cycle generated by $r^{1/2}(\lambda)$ and this operator are unitarily equivalent to the cycle generated by g and the operator σ^* induced by σ. Consequently g is an $L^2(m)$ coboundary. By Proposition 4.2 we conclude that g is an F_θ coboundary, which proves the first statement in the proposition.

The central limit theorem states that if $f,g \in F_\theta$, $\sigma_f^2(g) > 0$ and $\int g\,dm = 0$ where m is the equilibrium state for f then

$$G_n(y) = m\{x : g^n/\sqrt{n} < y\} \to N(y)$$

where N is the normal distribution with variance $\sigma^2 = \sigma_f^2(g)$ i.e.

$$\frac{dN(y)}{dy} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2} = n(y).$$

For a proof based on the Ruelle operator cf. [82]. We shall also provide a proof (taken from [23]) but with an approximation estimate of the order $O\left(\frac{1}{\sqrt{n}}\right)$ and for generic g of order $o\left(\frac{1}{\sqrt{n}}\right)$. Related work appears in [50], [52]. (See also [35**] and [76*].)

We shall say that g is *generic* if the equation $F(\sigma x) = e^{itg}F(x)$ (with F measurable or $F \in F_\theta$) has only the trivial solution $t = 0$, F constant. When $f, g \in F_\theta^+$, and f is normalised we have seen in Proposition 4.4 that this condition is equivalent to the requirement that L_{f+itg} has spectral radius less than 1 for all $t \neq 0$.

THEOREM 4.13. *Let* $f, g \in F_\theta$ *and suppose* $\sigma_f^2(g) > 0$ *and* $\int g dm = 0$ *where* m *is the equilibrium state of* f. *Then*

$$G_n(y) = N(y) + O\left(\frac{1}{\sqrt{n}}\right)$$

uniformly in y, *and if* g *is generic then*

$$G_n(y) = N(y) + \frac{P'''(0)}{6\sqrt{n}}\left(1 - \frac{y^2}{\sigma^2}\right)e^{-y^2/2\sigma^2} + o\left(\frac{1}{\sqrt{n}}\right).$$

PROOF. Since we are not considering asymptotics of a higher order than $o\left(\frac{1}{\sqrt{n}}\right)$ we are entitled to modify f,g by the additional of coboundaries. In other words there is no loss in generality in assuming that $f, g \in F_\theta^+$ and that f is normalised.

The proof of the Central Limit Theorem is based on a number of estimates. First we write

$$1 = w(s) + sv(s) \tag{4.2}$$

where $w(s) = Q(s)1$ and $sv(s) = (I-Q(s))1$ (since $w(0) = 1$). As $sv(s) \in (I-Q(s))F_\theta^+$ we have $v(s) \in (I-Q(s))F_\theta^+$ and therefore $-w'(0) = v(0) \in (I-Q(0))F_\theta^+$. From this we see that $\int v(0)dm = 0$ and conclude that

$$\int v(s)dm = s\psi(s), \text{ with } \psi(s) \text{ analytic.} \tag{4.3}$$

We shall also need the expression

$$P(f+sg) = \frac{\sigma^2 s^2}{2} + \frac{P'''(0)}{6} s^3 + s^4 \varphi(s) \tag{4.4}$$

where $\sigma^2 = \sigma_f^2(g) > 0$ and $\varphi(s)$ is analytic, which follows from the fact that $P(f) = P'(0) = 0$.

Using (4.4) and the elementary inequality $| e^{z+ib} - (1+ib) | \leq |z| e^{|z|} + \frac{b^2}{2}$ (for real b and complex z) we get

$$| e^{nP(f+itg/\sqrt{n})} - e^{-\sigma^2 t^2/2}\left(1 - \frac{it^3 P'''(0)}{6\sqrt{n}}\right) | \tag{4.5}$$

$$\leq e^{-\sigma^2 t^2/2}\left(\frac{t^4}{n} | \varphi\left(\frac{it}{\sqrt{n}}\right) | e^{\frac{t^4}{n}|\varphi|} + \frac{t^6}{72n} |P'''(0)|^2\right) = O\left(\frac{1}{n}\right)$$

uniformly for $|t| \leq \varepsilon\sqrt{n}$ if ε is chosen small enough. The implied constant then depends only on ε.

From these estimates, and using the fact that

$$\chi_n(t) \equiv \int e^{itg^n/\sqrt{n}}\, dm = \int L^n_{f+itg/\sqrt{n}}\ 1dm$$

one obtains:

For suitably small $\varepsilon > 0$ we have

$$\int_0^{\varepsilon\sqrt{n}} \frac{1}{t} | \chi_n(t) - e^{-\sigma^2 t^2/2}\left(1 - \frac{it^3 P'''(0)}{6\sqrt{n}}\right) | \, dt = O\left(\frac{1}{n}\right) \tag{4.6}$$

where the implied constant depends only on ε.

We define the distribution function

$$G_n(y) = m\{x \in X : g^n(x)/\sqrt{n} < y\}$$

whose Fourier transform is $\chi_n(t) = \int e^{itg^n/\sqrt{n}}\, dm$. If $G(y)$ has a continuous derivative $G'(y)$ whose Fourier transform $\gamma(t)$ satisfies $\gamma(0) = 1, \gamma'(0) = 0$ and if $G(-\infty) = 0, G(\infty) = 1$ then a well-known inequality (cf. [31]) asserts that

$$|G_n(y) - G(y)| \le \frac{1}{2\pi} \int_0^T \frac{1}{t} |\chi_n(t) - \gamma(t)|\, dt + \frac{24M}{\pi T} \tag{4.7}$$

where M is the maximum of $G'(y)$, and T is *any* positive number.

Applying (4.7), with $G(y) = N(y)$, the normal distribution with variance σ^2, so that $\gamma(t) = e^{-\sigma^2 t^2/2}$ we get

$$|G_n(y) - N(y)| \le O\left(\frac{1}{\sqrt{n}}\right) + \frac{24M}{\pi\varepsilon\sqrt{n}} = O\left(\frac{1}{\sqrt{n}}\right)$$

which proves the central limit theorem.

Actually (4.6) enables us to prove the C.L.T. with a $o\left(\frac{1}{\sqrt{n}}\right)$ asymptotic when g satisfies the generic condition. For in this case we define

$$G(y) = N(y) + \frac{P'''(0)}{6\sqrt{n}}\left(1 - \frac{y^2}{\sigma^2}\right)e^{-y^2/2\sigma^2}$$

which has a derivative $G'(y)$ whose Fourier transform is

$$\gamma(t) = e^{-\sigma^2 t^2/2}\left(1 - \frac{(it)^3 P'''(0)}{6\sqrt{n}}\right)$$

and (4.6) shows that

$$\int_0^{\varepsilon\sqrt{n}} \frac{1}{t} \mid \chi_n(t) - \gamma(t) \mid dt = O\left(\frac{1}{n}\right).$$

However, we shall see that

$$\int_{\varepsilon\sqrt{n}}^{\alpha\sqrt{n}} \frac{1}{t} \mid \chi_n(t) - \gamma(t) \mid dt \to 0$$

at an exponential rate whenever $\alpha > \varepsilon$.

From this it will follow that

$$|G_n(y) - G(y)| \leq O\left(\frac{1}{n}\right) + \frac{24M}{\pi\alpha\sqrt{n}}$$

for *all* $\alpha > \varepsilon$ which will prove the theorem.

It remains, then, to prove $\int_{\varepsilon\sqrt{n}}^{\alpha\sqrt{n}} \frac{1}{t} |\chi_n(t) - \gamma(t)| \, dt \to 0$ exponentially fast or, equivalently, that $\int_{\varepsilon\sqrt{n}}^{\alpha\sqrt{n}} \frac{1}{t} |\chi_n(t)| \, dt \to 0$ exponentially fast.

But the latter integral equals

$$\int_{\varepsilon}^{\alpha} \left| \int \exp(iyg^n) \, dm \right| \frac{dy}{y} = \int_{\varepsilon}^{\alpha} \left| \int L^n_{f+iyg} 1 \, dm \right| \frac{dy}{y}$$

and since g is generic L_{f+iyg} has spectral radius less than 1 for all real $y \neq 0$. So the proof of the theorem is complete.

Notes

The main result of this chapter is Theorem 4.5. This should be viewed as the complex analogue of the Ruelle operator theorem (Theorem 2.2). This theorem appeared in an article by Pollicott [71], developing a restricted version which appeared in the article of Parry-Pollicott [66]. However, the proof we give here differs from the original.

The statement and proof of Weilandt's theorem for complex matrices appears in Gantmacher's book [35].

A comprehensive account of (analytic) perturbation theory appears in Kato's well-known book [44]. A very nice account, which suffices for our needs, appears in the notes of Bhatia and Parthasarathy [8]. The application of this theory to the proof of analyticity of pressure is due to Ruelle [82].

Expressions for the first and second derivatives of pressure can be found in Ruelle's book [82] as exercises. (For the special case of locally constant functions these computations were independently derived (but later) by Parry-Tuncel [64].)

Proposition 4.12 appears in Ruelle's book [82].

For a brief account of Herglotz's theorem and the spectral density we refer the reader to the appendix in [62].

The central limit theorem has a very long history: recent contributions in the context of hyperbolic systems include those by Sinai [91], Ratner [74], [75], Denker-Phillip [27]. The basic idea of using the Ruelle operator appears in [82] (cf. also the articles of Keller [50] and Lalley [52].) The account given here follows [23] which is close to that of J. Rousseau-Egelé [76*] and Guivarc'h and Hardy [35**] - as the referee pointed out.

CHAPTER 5

PERIODIC POINTS AND ZETA FUNCTIONS

A convenient way of recording the number of periodic points of σ is through the zeta function defined formally by:

$$\zeta(z) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \nu_n \qquad (z \in \mathbb{C})$$

where $\nu_n = \sum_{\mathrm{Fix}_n} 1 = \mathrm{Card}\,(\mathrm{Fix}_n)$ and $\mathrm{Fix}_n = \{x : \sigma^n x = x\}$.

More generally, we can 'weight' periodic orbits by some function $f \in F_\theta$ and define

$$\zeta(z,f) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \sum_{\mathrm{Fix}_n} e^{f^n(x)} .$$

In view of Proposition 2.2 we can choose $g \in F_{\theta^{\frac{1}{2}}}(X^+)$ with $f \sim g$ and observe that $\zeta(z,f) = \zeta(z,g)$ since $f^n(x) = g^n(x)$ whenever $\sigma^n x = x$. Thus we may freely suppose that f is a function depending only on future coordinates.

The following proposition gives information on where $\zeta(z,f)$ is well-defined as a complex function.

PROPOSITION 5.1. *The radius of convergence of* $\sum_{n=1}^{\infty} \frac{z^n}{n} \sum_{x\in \mathrm{Fix}_n} e^{f^n(x)}$ *is* $e^{-P(f)}$ *when* f *is real. In particular, the radius of convergence of* $\sum_{n=1}^{\infty} \frac{z^n}{n} \nu_n$ *is* e^{-h}, *where* $h = P(0)$ *is the topological entropy of* σ.

PROOF. We shall actually prove a slightly stronger result, namely that:

$$\frac{1}{n} \log \sum_{\mathrm{Fix}_n} e^{f^n(x)} \to P(f) \text{ as } n \to +\infty .$$

For any given $\varepsilon > 0$ and $f \in F_\theta$ we can choose a function g of finitely many coordinates with $|f-g|_\infty < \varepsilon$. We may assume that g depends only on $x_0,\dots,x_\ell$, otherwise we replace g by $g \circ \sigma^r$, for sufficiently large r. Next we can replace words of length ℓ by symbols, if necessary, to assume that g is a function of x_0, x_1.

Let A_g be the matrix with entries $A_g(i,j) = A(i,j)\, e^{g(i,j)}$. Then

$$\sum_{\mathrm{Fix}_n} e^{g^n(x)} = \sum_{x_0\dots x_{n-1}x_0} e^{g(x_0,x_1)} e^{g(x_1,x_2)} \dots e^{g(x_{n-1},x_0)} = \mathrm{Trace}\ A_g^n = e^{nP(g)} + \lambda_2^n + \dots + \lambda_k^n$$

where $e^{P(g)}, \lambda_2,\dots,\lambda_k$ are the eigenvalues of A_g and $|\lambda_i| < e^{P(g)}$, $i = 1,\dots,k$ (if we assume that A is aperiodic).

Clearly for such g, $\frac{1}{n}\log \sum_{Fix_n} e^{g^n(x)} \to P(g)$, as claimed.

Since $e^{-n\varepsilon} \sum_{Fix_n} e^{g^n(x)} \leq \sum_{Fix_n} e^{f^n(x)} \leq e^{n\varepsilon} \sum_{Fix_n} e^{g^n(x)}$

we have $-\varepsilon + P(g) \leq \underline{\lim}\, \frac{1}{n}\log \sum_{Fix_n} e^{f^n(x)} \leq \overline{\lim}\, \frac{1}{n}\log \sum_{Fix_n} e^{f^n(x)} \leq \varepsilon + P(g)$.

Finally, since P is Lipschitz, with Lipschitz constant 1, we see that $-2\varepsilon + P(f) \leq -\varepsilon + P(g)$ and $\varepsilon + P(g) \leq 2\varepsilon + P(f)$. Since $\varepsilon > 0$ is arbitrary the result follows.

When f is a function of finitely many coordinates we can always assume $f(x) = f(x_0, x_1)$, after recoding, if necessary. We can then write:

$$\zeta(z,f) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \sum_{Fix_n} e^{f^n(x)}$$

$$= \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \text{ Trace } A_f^n = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \left(e^{nP(f)} + \lambda_2^n + \cdots + \lambda_k^n\right)$$

(as in the proof of the above proposition).

Hence for $|z| < e^{-P(f)}$,

$$\zeta(z,f) = \exp - \log\,[(1-ze^{P(f)})\ldots(1-z\lambda_k)]$$

$$= \frac{1}{\det(I-zA_f)}\ .$$

As a special case we see that for $|z| < e^{-h}$,

$$\zeta(z) = \frac{1}{\det(I-zA)}\ .$$

In each case, the closed form on the right-hand side gives a meromorphic extension to $\mathbb{C}$ of the zeta function.

Before considering the meromorphic extensions of more general zeta functions it is useful to make some observations that will prove useful later.

PROPOSITION 5.2. *Let* $f \in F_\theta$ *and suppose* $f^n(x) \in \mathbb{Z}a$, *for some real constant* a, *whenever* $\sigma^n x = x$. *Then* $f \sim f'$ *where* f' *takes values only in* $\mathbb{Z}a$.

PROOF. By taking a cohomologous function if necessary, we can assume $f \in F_{\theta^{\frac{1}{2}}}^{+}$ (cf. Proposition 1.2). If $a > 0$, say, then by adding a multiple of a we can assume that $f > 0$. Next we can multiply by a suitable *negative* constant so that $f < 0$ and $P(f) = 0$. A further addition of a coboundary allows us to assume that $f \in F_{\theta^{\frac{1}{2}}}^{+}$ and L_f is normalised (cf. Theorem 2.2). By Proposition 4.9 it is easy to see that $\rho\left(L_{f+\frac{2\pi i}{a}f}\right) = e^{P\left(\mathcal{R}\left(f+\frac{2\pi i}{a}f\right)\right)} = 1$. Thus by the comments preceding Theorem 4.5 we can conclude that $w(\sigma x) = w(x)\, e^{\frac{2\pi i}{a}f}$, for some $w \in F_\theta^{+}$ with $|w| = 1$. With $w(x) = e^{2\pi i v(x)}$ where v is real valued and continuous we have $av \circ \sigma = av + f + aM$, where M is integer valued. This completes the proof if $a \neq 0$. If $a = 0$ then, for instance, $f^n(x) = 0$ whenever $\sigma^n x = x$ and we have $P(tf) = P(0)$ for all $t \in \mathbb{R}$. It follows from Proposition 4.10 that $\frac{dP(tf)}{dt}\Big|_{t=1} = \int f dm = 0$, where m is the equilibrium state for f. Thus we can deduce that $P(f) = h_m(\sigma) + \int f dm = h_m(\sigma) = P(0) = h_{m_\circ}(\sigma)$, where $m_\circ$ is the measure of maximal entropy. In particular, f and 0 have precisely the same equilibrium state and thus by Proposition 3.6 we deduce that $f \sim 0$. This concludes the proof. (See also Proposition 3.7.)

Our approach to the meromorphic extension of zeta functions makes essential use of the following simple lemma on the spectra of Ruelle operators.

LEMMA 5.3. *Let* $f \in F_\theta^+$ *be in the domain of* P *and suppose* $L_f w(f) = e^{P(f)} w(f)$, *where* $w(f)$ *is nowhere vanishing (by reducing the domain of* P, *if necessary). Then there exists* $\varepsilon > 0$ *and* $N > 0$ *such that for all* $g \in D_\varepsilon(f) = \{g : \|f-g\|_\theta < \varepsilon\}$ *and all* $n \geq N$ *there exist* g_n, w_n, *functions of* $x_0, x_1, \ldots, x_n$ *such that*

$$|g_n - g|_\infty \leq K\theta^n, \ |w_n - w| \leq K\theta^n$$

for some constant K *(depending only on* f) *where*

$$L_{g_n} w_n = e^{P(g)} w_n .$$

PROOF. By Theorem 4.5 and the perturbation theorem, we can choose a $\| \ \|_\theta$ neighbourhood of f on which P is well defined and for which $L_g w = e^{P(g)} w$, where $w = w(g)$ has an analytic dependence on $g \in F_\theta^+$. We can also suppose that $|w(g)| \geq c > 0$.

We begin by choosing functions g'_n, w'_n depending only on $x_0, \ldots, x_n$ such that $|g - g'_n|_\infty \leq |g|_\theta \, \theta^n$ and $|w - w'_n|_\infty \leq |w|_\theta \, \theta^n$. It is easy to see that

$$|L_g w - L_{g'_n} w'_n|_\infty \leq C_1 \theta^n$$

(for some constant C_1 depending only on f).

Define $w_n = e^{-P(g)} L_{g'_n} w'_n$ and then it is simple to show that

$$\left|\frac{w_{n'}}{w_n} - 1\right| = \left|\frac{w_{n'} - w_n}{w_n}\right| = \left|\frac{(L_{g'_n} w'_n - L_g w) + e^{P(g)}(w - w'_n)}{L_g w - (L_g w - L_{g'_n} w'_n)}\right|$$

$$\leq \frac{|L_{g'} w'_n - L_g w| + e^{|P(g)|}|w - w'_n|_\infty}{e^{-|P(g)|}c - |L_g w - L_{g'_n} w'_n|_\infty} \leq C_2 \theta^n ,$$

for some constant C_2 depending only on f and for n sufficiently large. Writing $\frac{w_{n'}}{w_n} = e^{a_n + ib_n}$ with $-\pi \leq b_n < \pi$ we see that $|a_n + ib_n|_\infty \leq C_3 \theta^n$ (for some constant C_3 depending only on f).

By construction we have, $e^{P(g)} w_n = L_{g'_n} w'_n = L_{g'_n}((w'_n / w_n)\, w_n)$

$$= L_{g'_n + a_n + ib_n}(w_n) .$$

Thus with $g_n = g'_n + a_n + ib_n$ we have $L_{g_n} w_n = e^{P(g)} w_n$ and $|g - g_n|_\infty \leq K\theta^n$, $|w - w_n| \leq K\theta^n$, K constant and n sufficiently large.

COROLLARY 5.3.1. *With the notation of the lemma we have*

$$\|g - g_n\|_{\theta^{1/2}} \leq K' \theta^{n/2} , \|w - w_n\|_{\theta^{1/2}} \leq K' \theta^{n/2} , \text{ where } K' \text{ depends only on } f.$$

PROOF. We shall only deal with the first inequality, the second being somewhat similar. By the lemma we know that $\|g - g_n\|_\infty \leq K\theta^n$ so that

$$\frac{\mathrm{var}_k(g-g_n)}{\theta^{k/2}} \le \frac{2|g-g_n|_\infty}{\theta^{n/2}} \le 2K\theta^{n/2},$$

for $k = 0,1,2,\ldots,n$. Whereas

$$\frac{\mathrm{var}_k(g-g_n)}{\theta^{k/2}} \le \frac{\mathrm{var}_k\, g}{\theta^{k/2}} \le |g|_\theta\, \theta^{k/2} \quad \text{for } k > n\,.$$

So we may choose $K' \ge 2K$ and then $|g|_\theta \le K'$ for all g in an appropriate neighbourhood of f. (This is similar in spirit to Proposition 1.3.)

To analyse the domains of zeta functions it is appropriate to examine the series

$$Z(g) = \sum_{n=1}^{\infty} \frac{1}{n} \sum_{\mathrm{Fix}_n} e^{g^n(x)},$$

for g in a neighbourhood of $f \in F_\theta^+$, where $P\mathcal{R}(f) \le 0$. Our first result is straightforward.

THEOREM 5.4. *If* $f \in F_\theta^+$ *and* $P\,\mathcal{R}(f) < 0$ *then there exists* $\varepsilon > 0$ *such that* $Z(g)$ *converges absolutely in* $D_\varepsilon(f)$.

PROOF. This follows from

$$\Big|\sum_{\mathrm{Fix}_n} e^{g^n(x)}\Big|^{1/n} \le \Big(\sum_{\mathrm{Fix}_n} e^{\mathcal{R}g^n(x)}\Big)^{1/n} \to e^{P\mathcal{R}(g)} \le e^{P\mathcal{R}(f)+\varepsilon} < 1\ ,$$

as long as $P(\mathcal{R}f) < -\varepsilon$ and $g \in D_\varepsilon(f)$.

The next result is similar to the above, except that we must deal with the terms $\sum_{\mathrm{Fix}_n} e^{g^n(x)}$ in a slightly more delicate way.

THEOREM 5.5. *Let* $f \in F_\theta^+$ *then* $\rho(L_f) \le e^{P(\mathcal{R}f)}$ *(Theorem 4.5). Assume that* $P(\mathcal{R}f) = 0$.

(i) *If* $\rho(L_f) < 1$ *then there exists* $\varepsilon > 0$ *such that* $Z(g) = \sum_{n=1}^{\infty} \frac{1}{n} \sum_{\mathrm{Fix}_n} e^{g^n(x)}$ *converges absolutely in* $D_\varepsilon(f)$.

(ii) *If* $\rho(L_f) = 1$ *or equivalently* $f \sim \mathcal{R}f + ia + 2\pi iM$ *with* $M \in C(X, \mathbb{Z})$), *then there exists* $\varepsilon > 0$ *such that* $Z_1(g) = \sum_{n=1}^{\infty} \frac{1}{n} \Big(\sum_{\mathrm{Fix}_n} e^{g^n(x)} - e^{nP(g)}\Big)$ *converges absolutely in* $D_\varepsilon(f)$.

PROOF. The proof of part (i) is simpler than the proof of part (ii). We shall give the proof of part (ii) in detail and indicate the modifications needed for the first part.

We assume that P is defined for $f \in F_\theta^+$ $(0 < \theta < 1)$ and suppose $\mathcal{R}P(f) = P\mathcal{R}(f) = 0$. By perturbation theory there exist $\eta, \varepsilon > 0$ such that P is defined for $g \in D'_\varepsilon$ where $D'_\varepsilon = \{g \in F_{\theta^{\frac{1}{2}}}^+ : \|g-f\|_{\theta^{1/2}} \le 2\varepsilon\}$ and $|e^{P(g)} - e^{P(f)}| < \eta$, and the rest of the spectrum of $L_g : F_{\theta^{\frac{1}{2}}}^+ \to F_{\theta^{\frac{1}{2}}}^+$ is contained in $\{z \mid |z| < 1 - 2\eta\}$.

We can also assume that $\varepsilon > 0$ is chosen in accordance with the lemma, i.e. we have functions g_n, w_n of $x_0, x_1, \ldots, x_n$ satisfying $L_{g_n} w_n = e^{P(g)} w_n$, $L_g w = e^{P(g)} w$ and $\|g-g_n\|_{\theta^{1/2}} \le K'\theta^{n/2}$, $\|w-w_n\|_{\theta^{1/2}} \le K'\theta^{n/2}$ for all $n \ge N$, whenever $\|g-f\|_\theta \le \varepsilon$.

Let $0 < \alpha < 1$ (to be specified later) and denote $\nu = [n\alpha]$. Now consider $n \ge \dfrac{N+1}{\alpha}$, so that in particular $\nu \ge N$ and $n \ge N$. If $g \in D_\varepsilon(f)$ then

$$\left| \sum_{\mathrm{Fix}_n} e^{g^n(x)} - e^{nP(g)} \right| \le \left| \sum_{\mathrm{Fix}_n} e^{g^n(x)} - \sum_{\mathrm{Fix}_n} e^{g_\nu^n(x)} \right| + \left| \sum_{\mathrm{Fix}_n} e^{g_\nu^n(x)} - e^{nP(g_\nu)} \right|$$

(observing that $P(g) = P(g_\nu)$).

Moreover, $\sum_{\mathrm{Fix}_n} \left| e^{g^n(x)} - e^{g_\nu^n(x)} \right| \le \sum_{\mathrm{Fix}_n} nK'\theta^\nu \, e^{\mathcal{R}g^n(x) + nK'\theta^\nu}$, so that

$$\overline{\lim} \sum_{\mathrm{Fix}_n} \left| \left(e^{g^n(x)} - e^{g_\nu^n(x)} \right)^{1/n} \right| \le \theta^\alpha e^{P(\mathcal{R}(f))+\varepsilon} = \theta^\alpha e^\varepsilon < 1$$

providing $\varepsilon > 0$ is sufficiently small depending on α.

On the other hand consider the finite dimensional operator L_ν which is the restriction of L_{g_ν} to functions of $\nu + 1$ coordinates. Each eigenvalue of L_ν is in the spectrum of $L_{g_\nu} : F^+_{\theta^{\frac{1}{2}}} \to F^+_{\theta^{\frac{1}{2}}}$ and since $e^{P(g)}$ is an eigenvalue of L_ν and $P(g)$ is close to $P(f)$ we may suppose that the rest of the spectrum of L_ν consists of eigenvalues of modulus less than $1 - 2\eta$.

However, $\sum_{\mathrm{Fix}_n} e^{g_\nu^n(x)} = \text{Trace } L_\nu^n$, so that $\sum_{\mathrm{Fix}_n} e^{g_\nu^n(x)} - e^{nP(g)}$ is the sum of the n'th powers of at most k^ν numbers of modulus less than $(1 - 2\eta)^n$, i.e.

$$\overline{\lim} \left| \sum_{\mathrm{Fix}_n} e^{g_\nu^n(x)} - e^{nP(g)} \right|^{1/n} \le \overline{\lim} \left(k^\nu (1-2\eta)^n\right)^{1/n} = k^\alpha (1-2\eta) .$$

(Here, k is the dimension of the incidence matrix A.) We now assume that α has been chosen sufficiently small that $k^\alpha (1-2\eta) < 1$, and $\varepsilon = \varepsilon(\alpha)$ satisfies $\theta^\alpha e^\varepsilon < 1$.

We have therefore shown that $\sum_{n=1}^{\infty} \frac{1}{n} \left(\sum_{\mathrm{Fix}_n} e^{g^n(x)} - e^{nP(g)}\right)$ converges uniformly in $D_\varepsilon(f)$. This completes the proof of part (ii).

The proof of part (i) is similar, except that in this case one bounds all eigenvalues away from the unit circle using the upper semi-continuity of $f \mapsto L_f \mapsto \rho(L_f)$ i.e. one chooses $\varepsilon > 0$ so that for $g \in D_\varepsilon(f)$ we have $\rho(L_g) < 1$. A similar approximation argument is used, except that the term $e^{nP(g)}$ does not appear.

The above theorem is the crux of our analysis of the zeta functions $\zeta(f) = \exp Z(f)$. We present below the result in its final form.

THEOREM 5.6. (Extension Theorem) *Let* $f \in F_\theta^+$ *and assume* $P(\mathcal{R}f) = 0$, *so that* $\rho(L_f) \le 1$.

(a) *If* $\rho(L_f) < 1$ *then there exists* $\varepsilon > 0$ *such that* $\zeta(g) = \exp Z(g)$ *is nowhere zero and analytic in* $D_\varepsilon(f)$.

(b) *If* $\rho(L_f) = 1$ *then* ζ *can be extended to a nowhere zero analytic function in* $D_\varepsilon(f)$ *if* L_f *does not have* 1 *as an eigenvalue, by defining*

$$\zeta(g) = \frac{\exp Z_1(g)}{1 - e^{P(g)}}.$$

(c) *If* $\rho(L_f) = 1$ *and* L_f *has* 1 *as an eigenvalue then* ζ *can be extended to a nowhere zero analytic function in* $D_\varepsilon(f) - \{g : P(g) = 0\}$ *by defining*

$$\zeta(g) = \frac{\exp Z_1(g)}{1 - e^{P(g)}} .$$

PROOF. By theorem 5.5 we know that in each case we have defined a non-zero analytic function. To show that these are actually the meromorphic extensions of $\exp Z(g)$ we need to check that $\exp Z(g) = \dfrac{\exp Z_1(g)}{1-e^{P(g)}}$ for $\mathcal{R}P(g) < 0$, i.e. $|e^{P(g)}| < 1$. In this range each series is uniformly convergent and so this becomes a simple manipulation.

The above theorem is far more general than we shall need for our applications. For the analysis of hyperbolic flows and suspended flows we shall need the following example:

EXAMPLE. Fix $f \in F_\theta^+$ with $f > 0$ and $P(-f) = 0$ and define

$$\zeta_{-f}(s) = \zeta(-sf) = \exp \sum_{n=1}^{\infty} \frac{1}{n} \sum_{\mathrm{Fix}_n} e^{-sf^n(x)} ,$$

where $s \in \mathbb{C}$. We have seen that this is well defined, non-zero and analytic where $P(\mathcal{R}(-sf)) < 0$, i.e. $\mathcal{R}(s) > 1$. When $s_0 = 1+it_0$, ζ_{-f} has a non-zero analytic extension to a neighbourhood of s_0 when $L_{-(1+it_0)f}$ does *not* have 1 as an eigenvalue. When 1 *is* an eigenvalue we know that $\zeta_{-f}(s)$ has a non-zero analytic extension to $D_\varepsilon(1+it_0) - \{s : P(-sf) = 0\}$, where

$$D_\varepsilon(1+it_0) = \{s \in \mathbb{C} : |s-(1+it_0)| < \varepsilon\},$$

for sufficiently small $\varepsilon > 0$. To show that $\zeta_{-f}(s)$ has a non-zero analytic extension to $D_\varepsilon(1+it_0) - \{1+it_0\}$, for sufficiently small $\varepsilon > 0$, it suffices to show that $1 + it_0$ cannot be a point of accumulation of $\{s : P(-sf) = 0\}$. If $P(-sf) = 0$ for infinitely many s accumulating to $1 + it_0$ then by analyticity $P(-sf) \equiv 0$ in a neighbourhood of $1 + it_0$. Hence $L_{-(1+it)f}\, w_t = w_t$, for t near t_0 so $w_t \circ \sigma = e^{-itf} w_t$ with $|w_t| \equiv 1$. We can write $w_t = e^{2\pi i v_t}$, with v_t continuous and real-valued, then $v_t \circ \sigma = -\frac{tf}{2\pi} + v_t + M_t$, M_t integer valued. If $\int f d\mu \neq 0$, where μ is the equilibrium state of $-f$ then we arrive at a contradiction since $\frac{t}{2\pi}\int f d\mu = \int M_t d\mu$ can only take a countable number of values. If $\int f d\mu = 0$ then $P(-f) = 0 = h_\mu(\sigma) - \int f d\mu = h_\mu(\sigma)$. However, the entropy $h_\mu(\sigma)$ cannot be zero, giving the required contradiction.

Notes

The zeta-function $\zeta(z)$ for diffeomorphisms is discussed in the work of Artin-Mazur [7] and Smale [95]. For subshifts of finite type calculations were made by Bowen and Lanford [9]

The weighted zeta-function $\zeta(z,f)$ is studied in the work of Ruelle [80] and Bowen [15].

Theorem 5.6 (b) was proved in an article by Pollicott [71], whereas Theorem 5.6 (c) comes from Parry's article [67] (containing also the constructions from Lemma 5.3 and Corollary 5.3.1).

Theorem 5.6 (a) is implicit in the work of Bowen [15] (cf. also Ruelle's article [80]).

Ruelle proved an earlier partial version of Theorem 5.6 (c), where the dependence on the second variable is restricted to one-dimension, i.e. $(z,s) \mapsto \zeta(z,sf)$ (cf. [80]). This result of Ruelle was extended by Parry-Pollicott in [66], to obtain the version described in the example at the end of the section. This was a preliminary version of Theorem 5.6.

Under certain analytic hypotheses Ruelle was able to obtain a meromorphic extension of $\zeta(z,f)$ to the entire complex plane [79]. Similarly, in certain smooth settings Tangerman has shown that $\zeta(z,f)$ extends to the entire plane [99]. However, because of the lack of smoothness of the stable manifold foliations these results are not immediately applicable to the context of hyperbolic flows except in exceptional cases. For geodesic flows associated to compact manifolds with constant negative sectional curvatures the associated zeta-function has a meromorphic extension to $\mathbb{C}$, using an approach of Selberg [85] (cf. [38]).

CHAPTER 6

PRIME ORBIT THEOREMS FOR SUSPENDED FLOWS

In this chapter we shall introduce suspended flows and associate to them a natural zeta function incorporating information about closed orbit periods. Information on the domains of these zeta functions can be deduced from the more general analysis in the previous chapter. We shall then explain the role of these zeta functions in deducing asymptotic formulae for closed orbit periods.

Let σ be the shift defined by A and let $f \in F_\theta(x)$ be strictly positive (with $0 < \theta < 1$). We define the *suspension space* (relative to f) as

$$X_f = \{(x,y) : x \in X, 0 \leq y \leq f(x)\}$$

with the identification $(x, f(x)) = (\sigma x, 0)$. An alternative definition is $X_f = X x \mathbb{R}/\mathbb{Z}$ where $\mathbb{Z}$ is the group of maps generated by $(x,y) \mapsto (\sigma x, y-f(x))$.

The *suspension flow* σ_f (relative to f) is defined as the "vertical" flow on X_f given by $\sigma_{f,t}(x,y) = (x, y+t)$, for small t. (This condition makes sense for $0 \leq y, y+t < f(x)$ and can be extended using the identifications.) Equivalently, $\sigma_{f,t}$ is the flow on $X x \mathbb{R}/\mathbb{Z}$ induced by maps $(x,y) \rightarrow (x,y+t)$. Clearly, these maps commute with the group $\mathbb{Z}$.

If f and f′ are cohomologous functions and f′ is also strictly positive it is easy to see that (X_f, σ_f) and $(X_{f'}, \sigma_{f'})$ are topologically conjugate, for it suffices to define a homeomorphism φ of $X x \mathbb{R}$ which commutes with the vertical flow and conjugates the two maps:

$$(x,y) \mapsto (\sigma x, y-f(x)) \text{ and } (x,y) \mapsto (\sigma x, y-f'(x)).$$

For example, when $f'(x) = f(x) + v(x) - v(\sigma x)$ it suffices to define $\varphi(x,y) = (x,y+v(x))$.

In view of the above observation we can replace f by a function depending only on future coordinates without making any essential change in the underlying flow. We shall do this whenever it proves convenient.

If μ is a σ-invariant probability measure then we define a $\sigma_{f,t}$-invariant probability measure μ_f by

$$\int_{X_f} F d\mu_f = \frac{\int_X \left(\int_0^{f(x)} F(x,y)\, dy \right) d\mu(x)}{\int_X f(x)\, d\mu(x)}.$$

In other words, μ_f is the normalisation of the measure on X_f obtained by taking the direct product of μ with Lebesgue measure on $\mathbb{R}$. In fact, it can be shown that *every* σ_f-invariant probability measure on X_f can be obtained in this way from a σ-invariant probability measure μ on X. Furthermore, it is easy to see that σ_f is ergodic with respect to μ_f if and only if σ is ergodic with respect to μ.

It has been shown by Abramov [2] (in somewhat more general circumstances) that

$$h(\sigma_{f,t}, \mu_f) = |t|\, h(\sigma_{f,1}, \mu_f)$$

from which one obtains the natural definition of the entropy of the flow as $h(\sigma_f, \mu_f) = h(\sigma_{f,1}, \mu_f)$ and the definition of *topological entropy* as $h = h(\sigma_f) = \sup_{\mu_f} h(\sigma_f, \mu_f)$.

The work of Abramov also relates the entropy of σ_f relative to μ_f to the entropy $h(\sigma,\mu)$ of σ relative to μ by

$$h(\sigma_f, \mu_f) = \frac{h(\sigma,\mu)}{\int f d\mu}.$$

The notions of pressure and of equilibrium state for σ_f are defined in analogy to the case of the shift σ. In particular, if $G \in C(X_f)$ we define the *pressure* by

$$P(G) = \sup \{h(\sigma_f, \mu_f) + \int G d\mu_f\}.$$

A σ_f-invariant probability measure μ_f is an *equilibrium state* of G if $P(G) = h(\sigma_f, \mu_f) + \int G d\mu_f$. We have used P to denote the pressure for continuous functions on both X and X_f; the context should make it clear which pressure is intended.

There is a simple relationship between the two pressures and two equilibrium states which we state as a proposition.

PROPOSITION 6.1. *Let* $G \in C(X_f)$ *be a real valued function and assume that* $g(x) = \int_0^{f(x)} G(x,t)dt \in F_\theta(X)$. *Then* $P(G) = c$ *where* c *is the unique real number such that* $P(g-cf) = 0$. *Moreover, if* m *is the unique equilibrium state of* $g-cf$ *then* m_f *is the unique equilibrium state of* G. *In particular, if* $G \equiv 0$, $(g \equiv 0)$ *then* $P(0) = h$, *the topological entropy of* σ_f, *and the measure of maximum entropy* m_f *for* σ_f *is unique, where* m *is the equilibrium state for* $-hf$.

PROOF. By the variational principle we can see that since $f > 0$ the map $c \mapsto P(g-cf)$ is strictly monotonic, with $\lim_{c \searrow -\infty} P(g-cf) = +\infty$, $\lim_{c \nearrow +\infty} P(g-cf) = -\infty$. In particular, there exists a unique constant c with $P(g-cf) = 0$. Consequently,

$$0 = h_m(\sigma) + \int (g-cf)dm \geq h_\mu(\sigma) + \int (g-cf)d\mu$$

for all σ-invariant probabilities μ with equality only when $\mu = m$.

Thus

$$c = \frac{h(m) + \int gdm}{\int fdm} \geq \frac{h(\mu) + \int gd\mu}{\int fd\mu}$$

i.e. $c = h(\sigma_f, m_f) + \int Gdm_f \geq h(\sigma_f, \mu_f) + \int Gd\mu_f$ with equality only when $\mu = m$ (or

$\mu_f = m_f)$. From the definitions we see that $c = P(G)$ and that equality only holds when $m_f = \mu_f$.

For the special case $G = 0$ we see that $h = h(\sigma_f, m_f)$ where m is the equilibrium state for $-P(G)f$. This completes the proof of the proposition.

The close relationship between the pressure functions allows one to deduce some results for $P : C(X_f) \to \mathbb{R}$ from results we showed earlier for $P : C(X) \to \mathbb{R}$. For example, if $W, G \in C(X_f)$, then $P(G+W \circ \sigma_{f,t_0} - W+a) = P(G)+a$, for a constant.

As remarked earlier we can freely assume that $f \in F_\theta^+$ (by moving to a conjugate flow). We can also interpret σ_f as a semi-flow on the space $X_f^+ = \{(x,y) : 0 \leq y \leq f(x) , x \in X^+\}$ with the usual identifications $(x, f(x)) \equiv (\sigma x, 0)$ and for $t > 0$, $\sigma_{f,t}(x,y) = (x,y+t)$, $0 < y$, $y+t < f(x)$.

Let m be the equilibrium state for any $u \in F_\theta^+$ and let m_f be the usual Lebesgue extension to X_f^+ , where we assume that $f \in F_\theta^+$.

PROPOSITION 6.2. *The following are equivalent:*

(i) σ_f *has an eigenfrequency* a *corresponding to an* $L^2(m_f)$ *function*

(ii) σ_f *has an eigenfrequency* a *corresponding to a continuous function*

(iii) $$w(\sigma x) = e^{iaf(x)}w(x)\,, \quad \textit{for some} \ w \in L^2(m)$$

(iv) $$w(\sigma x) = e^{iaf(x)}w(x)\,, \quad \textit{for some} \ w \in F_\theta^+$$

(v) $$L_{u+iaf}w = e^{P(u)}w\,, \textit{for some} \ w \in L^2(m)\,, \ \textit{or} \ F_\theta^+\,.$$

Furthermore, if any of these conditions hold then a *is isolated.*

PROOF. We already know from Chapter 4 that (iii), (iv), and (v) are equivalent.

Suppose (i) holds, and that $W\sigma_{f,t}(x,y) = e^{iat}W(x,y)$ a.e. (m_f) for $W \in L^2(m_f)$. In particular, $W(x,y+f(x)) = e^{iaf(x)}W(x,y)$ where $0 \le y \le \varepsilon$, for some sufficiently small $\varepsilon > 0$, a.e. (m_f) and so by Fubini's theorem there exists $y \in [0,\varepsilon]$ so that

$$W(\sigma x,y) = e^{iaf(x)}W(x,y) \ \text{ a.e. } (m)\,.$$

This shows (iii) to be true.

Assuming (iv) then we have $w(\sigma x) = e^{iaf(x)}w(x)$, $w \in F_\theta^+$. We can define $W(x,y) = w(x)e^{iay}$ and then

$$W(x,f(x)) = w(x)e^{iaf(x)} = w(\sigma x) = W(\sigma x,0)\,.$$

We conclude that $W(x,y)$ is well-defined and continuous on X_f. From the construction we have $W\sigma_{f,t}(x,y) = e^{iat}W(x,y)$. Thus (ii) is valid.

To see that any such a (occurring in the statement of the theorem) must be isolated we proceed as follows. Let λ be the least period of any closed orbit of σ_f. Thus $\sigma_{f,\lambda}(x) = x$ for some x and $w(x) = e^{ia\lambda}w(x)$. In particular, $a \in \left(\frac{2\pi}{\lambda}\right)\mathbb{Z}$, and we observe that the eigenfrequencies form a discrete subgroup of $\frac{2\pi}{\lambda}\mathbb{Z}$.

The flow σ_f is said to be *weak-mixing* if condition (i) (and therefore the other conditions) implies that $a = 0$ and the only L^2 eigenfunctions are the constant functions. (Conversely, σ_f is *not* weak-mixing if condition (i) is valid with a non-zero.)

We return to our consideration of zeta functions, and consider the implications of σ_f being weak-mixing or not. From our comments at the end of the previous chapter we know that if $f \in F_\theta^+$ with $f > 0$, $P(-f) = 0$ then $\zeta_{-f}(s)$ is nowhere zero analytic extension to $\mathcal{R}(s) = 1$ Furthermore, $\zeta_{-f}(s)$ has a nowhere zero analytic extension to $\mathcal{R}(s) = 1$ except for those $s_0 = 1+it_0$ where $w\sigma = e^{-it_0 f}w$ has a solution $w \in F_\theta^+$, $w \neq 0$. By Proposition 6.2 above the weak-mixing assumption implies that this equality only holds when $t_0 = 0$. Thus when σ_f is weak-mixing then $\zeta_{-f}(s)$ has a non-zero analytic extension to $\mathcal{R}(s) \geq 1$, except for $s = 1$.

Next we show that $s = 1$ is a simple pole for $\zeta_{-f}(s)$. We know that $\zeta_{-f}(s)(1-e^{P(-sf)})$ is non-zero and analytic in a neighbourhood of $s = 1$ so we need only observe that

$$\begin{aligned}\lim_{s\to 1} \frac{1-e^{P(-sf)}}{s-1} &= \lim_{s\to 0} \frac{1-e^{P(-f-sf)}}{s} \\ &= -\frac{dP(-(1+s)f)}{ds}\Big|_{s=0} \\ &= -\int f\,dm \neq 0\end{aligned}$$

where m is the equilibrium state for $-f$.

Summarising we have:

THEOREM 6.3. *If* σ_f *is weak-mixing with* $f \in F_\theta$, $P(-f) = 0$ *then* $\zeta_{-f}(s)$ *has a non-zero analytic extension to* $\mathcal{R}(s) \geq 1$, *except for a simple pole at* $s = 1$.

COROLLARY 6.3.1. *If* σ_f *is weak-mixing,* $f \in F_\theta$, $P(-f) = 0$ *then*

$\zeta'(s)/\zeta(s) = \dfrac{-1}{s-1} + \alpha(s)$ *where* $\alpha(s)$ *is analytic for* $\mathcal{R}(s) \geq 1$.

We can formulate a more general version of Theorem 6.3. Assume $f, g, k \in F_\theta$, where $f > 0$ *and* $c > 0$ is the unique real number satisfying $P(g-cf) = 0$. Of particular interest will be the special case $g(x) = \int_0^{f(x)} G(x,y)dy$, $k(x) = \int_0^{f(x)} K(x,y)dy$ and $c = P(G)$.

THEOREM 6.4. *If* σ_f *is weak-mixing then*

$$\zeta(s,z) = \exp \sum_{n=1}^{\infty} \frac{1}{n} \sum_{x \in \mathrm{Fix}_n} \exp\left(g^n - csf^n + zk^n\right)$$

is a nowhere zero analytic function for $\mathcal{R}(s) > 1$, z *in a neighbourhood of* 0 *(depending on* s), *with a nowhere zero analytic extension to* $\mathcal{R}(s) = 1$ $(s \neq 1)$, *for sufficiently small* $|z|$ *(depending on* s).

Furthermore, $\zeta(s,z)\,(1-e^{P(g-csf+zk)})$ *has a nowhere zero analytic extension to* $s = 1$, $|z|$ *sufficiently small (depending on* s).

PROOF. This is essentially a corollary of Theorem 5.6. We need only check that $s = 1+it_0$, $t_0 \neq 0$ does not occur as a singularity, for $|z|$ sufficiently small. If this were the case then $L_{g-(1+it_0)cf}$ would have $1 = e^{P(g-cf)}$ as an eigenvalue. But, as explained in Chapter 4, this would correspond to σ_f having an eigenfrequency $-t_0c$ (by Proposition 6.2). Since we are assuming σ_f to be weak mixing we conclude that $t_0 = 0$. For $s = 1$ we have, also by Theorem 5.6, $\zeta(s,z)(1-e^{P(g-csf+zk)})$ is nowhere zero and analytic in an (s,z) neighbourhood of $(1,0)$.

COROLLARY 6.4.1. *If* ζ_2 *denotes the derivative in the* z *coordinate then*

$$\frac{\zeta_2(s,0)}{\zeta(s,0)} = \frac{\int k\,dm}{c\int f\,dm}\frac{1}{(s-1)} + \alpha(s)$$

$$= \frac{\int K\,dm_f}{c(s-1)} + \alpha(s)$$

where $\alpha(s)$ *is an analytic function in* $\mathcal{R}(s) \geq 1$, *and* m, m_f *are the equilibrium states of* $g-cf, G$ *respectively (and* $c = P(G)$).

PROOF. Taking the logarithmic derivative (with respect to z) of $\zeta(s,z)$ gives

$$\frac{\partial}{\partial z}\log\zeta(s,z)\,\Big|_{z=0} = \frac{\zeta_2(s,0)}{\zeta(s,0)} = \frac{\frac{\partial P}{\partial z}(g-csf+zk)\,\Big|_{z=0}}{(e^{-P(g-csf)}-1)} + \alpha_0(s)$$

for s close to 1. Since

(i) $$\lim_{s\to 1}\frac{\partial}{\partial z}P(g-csf+zk)\,\Big|_{z=0} = \int k\,dm,$$

where m is the equilibrium state of $g-cf$, and

$$\text{(ii)} \quad \lim_{s\to 1} \frac{e^{-P(g-csf)}-1}{s-1} = \frac{\partial e^{-P(g-cfs)}}{\partial s}\Big|_{s=1} = \frac{-\partial P(g-cfs)}{\partial s}\Big|_{s=1} = \int cfdm$$

we see that $\dfrac{\zeta_2(s,0)}{\zeta(s,0)} = \dfrac{\int kdm / c\int fdm}{s-1} + \alpha(s) = \dfrac{\int Kdm_f}{c(s-1)} + \alpha(s)$

where α is analytic in $\mathcal{R}(s) \geq 1$. This completes the proof.

For the special case $G \equiv K \equiv 0$ we retrieve the zeta function of a single complex variable

$$\zeta_{-hf}(s) = \exp \sum_{n=1}^{\infty} \frac{1}{n} \sum_{x \in Fix_n} e^{-hf^n(x)s}$$

where $P(-hf) = 0$, i.e. $h = P(0)$, the topological entropy of the flow σ_f. We proceed to show that this zeta function can be expressed in other ways.

It is easy to see that there is a one-one correspondence between closed orbits $\{x,\sigma x,...,\sigma^{n-1}x\}$ for $\sigma : X \to X$ and closed orbits τ of the flow σ_f. Let $\lambda = \lambda(\tau)$ denote the least period of the closed orbit τ, i.e. $\sigma_{f,\lambda}(x,0) = (x,0)$, and it is the least such $\lambda > 0$. Then clearly $\lambda(\tau) = f(x) + f(\sigma x)+\cdots+f(\sigma^{n-1}x) = f^n(x)$, where n is the least positive integer such that $\sigma^n x = x$.

We can give an alternative expression for the zeta function as follows:

$$\zeta_{-hf}(s) = \exp \sum_{n=1}^{\infty} \frac{1}{n} \sum_{\mathrm{Fix}_n} e^{-hf^n(x)s}$$

$$= \exp \sum_{n=1}^{\infty} \frac{1}{n} \sum_{\substack{\sigma^n x = x \\ n \text{ least}}} \sum_{k=1}^{\infty} \frac{e^{-hkf^n(x)s}}{k}$$

$$= \exp \sum_{\tau} \sum_{n=1}^{\infty} \frac{e^{-hkf^n(x)s}}{k}$$

$$= \exp - \sum_{\tau} \log (1 - e^{-h\lambda(\tau)s})$$

$$= \prod_{\tau} (1 - e^{-h\lambda(\tau)s})^{-1} .$$

(These manipulations can be performed for $\mathcal{R}(s) > 1$ where $\zeta_{-hf}(s)$ converges.)

This final form is close to the classical definition of a zeta function, and provides the motivation for the terminology.

Before proceeding to prove the distribution result for closed orbits of the flow σ_f it is instructive to consider the analogous, and simpler, problem for the shift $\sigma : X \to X$.

As we observed in Chapter 5 we can express the zeta function for σ in the simple closed form $\zeta_A(z) = \dfrac{1}{\det (I-zA)}$. We can gain insight into the distribution of closed σ-orbit periods from the meromorphic domain of $\zeta_A(z)$.

Following the derivation above we can express ζ_A in the form

$$\zeta_A(z) = \prod_{\tau}(1-z^{\lambda(\tau)})^{-1},$$

where $\lambda(\tau)$ is the least period of a closed orbit τ for σ. Consequently,

$$\zeta_A(z) = \exp \sum_{\tau} \sum_{n=1}^{\infty} \frac{z^{\lambda(\tau)n}}{n}, \text{ for } |z| < 1/\beta$$

where $h = \log \beta$ is the topological entropy of σ and β is the maximal eigenvalue for A.

Since $\det(I-zA) = (1-\beta z) \prod_i (1-\lambda_i z)$, where the other eigenvalues of A satisfy $|\lambda_i| < \beta$ for an aperiodic matrix, we see that $\zeta_A(z)\,(1-\beta z)$ is non-zero and analytic in $\{z : |z| < e^{\varepsilon}/\beta\}$, for some $\varepsilon > 0$.

Thus $\quad \zeta'_A(z)/\zeta_A(z) = \sum_{\tau,n} \lambda(\tau) z^{\lambda(\tau)n-1} = \dfrac{\beta}{1-\beta z} + \alpha(z)$

where $\alpha(z)$ is analytic in $\{z \mid |z| < e^{\varepsilon}/\beta\}$.

For notational simplicity we denote by τ' a multiple closed orbit $\tau^n (n \geq 1)$.

(By a multiple closed orbit τ^n we mean a closed orbit τ of least period $\lambda(\tau)$ counted multiply as a closed orbit of period $n\lambda(\tau)$.) We write $\Lambda(\tau') = \lambda(\tau)$ and $\lambda(\tau') = n\lambda(\tau)$.

We therefore see that,

$$\frac{\zeta'_A(z)}{\zeta_A(z)} = \frac{1}{z}\sum_{\tau'} \Lambda(\tau') z^{\lambda(\tau')} = \frac{1}{z}\sum_{n=1}^{\infty} \beta^n z^n + \alpha(z).$$

Hence,

$$\sum_{n=1}^{\infty} z^{n-1}\Big(\sum_{\lambda(\tau')=n} \Lambda(\tau') - \beta^n\Big) = \alpha(z).$$

In particular, the radius of convergence of this series is at least e^{ε}/β. For a possibly smaller $\varepsilon > 0$, we deduce that

$$\frac{e^{n\varepsilon}}{\beta^n}\Big(\sum_{\lambda(\tau')=n} \Lambda(\tau') - \beta^n\Big) \to 0 \text{ as } n \to \infty.$$

For $x > 1$ we define

$$\psi(x) = \sum_{\lambda(\tau')\le x} \Lambda(\tau') = \sum_{n=1}^{x}\Big(\sum_{\lambda(\tau')=n} \Lambda(\tau') - \beta^n\Big) + \beta\,\frac{(\beta^x - 1)}{\beta - 1}$$

Clearly,

$$\left| \psi(x) - \frac{\beta^{x+1} - \beta}{\beta - 1} \right| \leq \frac{C\beta^x}{e^{\varepsilon x}}, \tag{6.1}$$

for some constant $C > 0$, by rearranging the above expressions.

Next we introduce $\pi'(x) = \sum_{\lambda(\tau) \leq x} 1$, which is simply the number of closed orbits whose period is less than or equal to x and we proceed to relate $\pi'(x)$ and $\psi(s)$:

(a) $$\psi(x) = \sum_{\lambda(\tau) \leq x} \lambda(\tau) \left[\frac{x}{\lambda(\tau)} \right] \leq x\pi'(x) ;$$

(b) If $x = \gamma y$ with $\gamma > 1$ then,

$$\pi'(x) = \pi'(y) + \sum_{y < \lambda(\tau) \leq x} 1$$

$$= \pi'(y) + \sum_{\lambda(\tau) \leq x} \frac{\lambda(\tau)}{y}$$

$$\leq \pi'(y) + \frac{\psi(x)}{y}$$

so that $\dfrac{x.\pi'(x)}{\beta^x} \le \dfrac{\gamma y.\pi'(y)}{\beta^{\gamma y}} + \dfrac{\gamma\psi(x)}{\beta^x}$.

We want to show that $\dfrac{\pi'(y)}{\beta^{\gamma' y}} \to 0$ whenever $\gamma' > 1$, from which (together with (6.1)) it will follow that $\overline{\lim}_{x\to+\infty} \dfrac{x.\pi'(x)}{\beta^x} \le \dfrac{\gamma.\beta}{\beta - 1}$

when $\gamma > \gamma' > 1$, for then we shall have (again by (6.1))

$$\overline{\lim}_{x\to+\infty} \frac{x.\pi'(x)}{\beta^x} \le \gamma\, \overline{\lim}_{y\to+\infty} \left(\frac{y}{\beta^{(\gamma-\gamma')y}}\right)\left(\frac{\pi'(y)}{\beta^{\gamma' y}}\right) + \gamma\, \overline{\lim}_{x\to+\infty} \frac{\psi(x)}{\beta^x} \le \frac{\gamma.\beta}{\beta - 1} \ . \tag{6.2}$$

Furthermore (6.1) and a) imply

$$\frac{\beta}{\beta - 1} \le \underline{\lim} \frac{x\pi'(x)}{\beta^x} \tag{6.3}$$

and since $\gamma > 1$ can be chosen arbitrarily close to one the two inequalities (6.2) and (6.3) will complete the proof of the following theorem.

THEOREM 6.5. $\pi'(x) \sim \dfrac{\beta}{\beta - 1} . \dfrac{\beta^x}{x}$ *as* $x \to \infty$.

The step required to complete the proof is the following:

LEMMA 6.6. *If* $\gamma > 1$ *then* $\dfrac{\pi'(y)}{\beta^{\gamma y}} \to 0$ *as* $y \to \infty$.

PROOF. We know that $\zeta_A(z)$ converges for $|z| < 1/\beta$ so

$$\begin{aligned}
\zeta_A(1/\beta^\gamma) &= \prod_\tau \left(1 - \frac{1}{\beta^{\gamma\lambda(\tau)}}\right)^{-1} \\
&\geq \prod_{\lambda(\tau)\leq y} \left(1 + \frac{1}{\beta^{\gamma\lambda(\tau)}}\right) \\
&\geq \left(1 + \frac{1}{\beta^{\gamma y}}\right)^{\pi'(y)} \\
&\geq 1 + \frac{\pi'(y)}{\beta^{\gamma y}} \geq \frac{\pi'(y)}{\beta^{\gamma y}} .
\end{aligned}$$

Thus $\dfrac{\pi'(y)}{\beta^{\gamma y}}$ is bounded for every $\gamma > 1$, and so we deduce $\dfrac{\pi'(y)}{\beta^{\gamma y}} \to 0$ for every $\gamma > 1$.

Having dealt with the closed orbits for σ we return to the analogous problem for the closed orbits of σ_f.

We shall assume that $f > 0$ and that σ_f is weak-mixing. We define for each closed orbit τ of σ_f the *norm* of τ to be $N(\tau) = e^{h\lambda(\tau)}$ where $h = P(0)$ is the topological entropy of σ_f.

Since $\zeta(s) = \zeta_{-hf}(s) = \exp \sum_{n=1}^{\infty} \sum_{\tau} \frac{1}{n} N(\tau)^{-sn}$

we can take the logarithmic derivative to get

$$\frac{\zeta'(s)}{\zeta(s)} = - \sum_{n=1}^{\infty} \sum_{\tau} \log N(\tau).N(\tau)^{-sn}$$

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

where α is analytic in $\mathcal{R}(s) \geq 1$.

Defining $S(x) = \sum_{N(\tau)^n \leq x} \log N(\tau) = \sum_{N(\tau) \leq x} n \log N(\tau)$ (where $n = \left[\frac{\log x}{\log N(\tau)} \right]$ i.e. the largest integer such that $N(\tau)^n \leq x$) we see that

$$\frac{\zeta'(s)}{\zeta(s)} = - \int_1^{\infty} x^{-s} \, dS(x) = \frac{-1}{s-1} + \alpha(s).$$

We shall need the following Tauberian theorem.

THEOREM 6.7. (*Ikehara-Wiener Tauberian Theorem.*) (See Appendix I.)

Assume that the Steiltjes integral $\int_1^\infty x^{-s}\, dS(x)$ *is defined and analytic in* $\mathcal{R}(s) > 1$ *with an analytic continuation to* $\mathcal{R}(s) \geq 1$ *except for a pole at* $s = 1$ *with residue* 1 *(i.e.* $\int_1^\infty x^{-s}\, dS(x) = \frac{1}{s-1} - \alpha(s)$, *with* $\alpha(s)$ *analytic in* $\mathcal{R}(s) \geq 1$*) then* $S(x) \sim x$ as $x \to +\infty$.

By Theorem 6.3 and Corollary 6.3.1 we know that in our case $\zeta'(s)/\zeta(s)$ has the correct analytic domain and, as a logarithic derivative, the pole at $s = 1$ must have residue 1. Hence $S(x) \sim x$ by applying the above theorem.

We conclude

$$x \sim \sum_{N(\tau)\leq x} \left[\frac{\log x}{\log N(\tau)}\right] \log N(\tau) \leq \log x.\pi(x)$$

where $\pi(x) = \sum_{N(\tau)\leq x} 1$. Hence $\underline{\lim}\, \frac{\pi(x)\log x}{x} \geq 1$.

We want to obtain an asymptotic upper bound on $\pi(x)$. Let $\gamma > 1$ and write $y = x^{1/\gamma} < x$. Then

$$\pi(x) = \pi(y) + \sum_{y<N(\tau)\leq x} 1 \leq \pi(y) + \sum_{N(\tau)\leq x} \frac{\log N(\tau)}{\log y}$$

$$\leq \pi(y) + \frac{1}{\log y} S(x)$$

$$= \pi(y) + \frac{\gamma}{\log x} S(x) .$$

So $\log x \dfrac{\pi(x)}{x} \leq \dfrac{\pi(y)}{y^{\gamma}} \gamma \log y + \dfrac{\gamma S(x)}{x}$.

We shall show that $\pi(y)/y^{\gamma} \to 0$ as $y \to +\infty$, whenever $\gamma > 1$ (or equivalently, $\pi(y)/y^{\gamma'}$ is bounded for each fixed $\gamma' > 1$). It will then follow that

$$\overline{\lim} \log x \frac{\pi(x)}{x} \leq \gamma , \text{ since } \overline{\lim_{y}} \frac{\pi(y) \log y}{y^{\gamma}} = \overline{\lim_{y}} \left(\frac{\pi(y)}{y^{\gamma'}}\right) \frac{\log y}{y^{\gamma-\gamma'}} = 0$$

and $\lim\limits_{x\to+\infty} \dfrac{S(x)}{x} = 1$ by the Tauberian theorem. Since $\gamma > 1$ was arbitrary we have $\overline{\lim} \log x \dfrac{\pi(x)}{x} \leq 1$. This is the desired asymptotic upper bound on $\pi(x)$. To complete the proof we need to establish the following lemma.

LEMMA 6.8. *If* $\gamma > 1$ *then* $\frac{\pi(y)}{y^\gamma}$ *is bounded for every large* y.

PROOF. For $\gamma > 1$ observe that

$$\zeta(\gamma) = \prod_\tau (1 - N(\tau)^{-\gamma})^{-1} \geq \prod_\tau (1 + N(\tau)^{-\gamma})$$

$$\geq \prod_{N(\tau)\leq y} (1+y^{-\gamma}) = (1+y^{-\gamma})^{\pi(y)} \geq \frac{\pi(y)}{y^\gamma}$$

Thus for all y large, $\frac{\pi(y)}{y^\gamma}$ is bounded by $\zeta(\gamma)$. Collecting together the above estimates gives our main result.

THEOREM 6.9. (*Prime Orbit Theorem*). *If* $f \in F_\theta$ *is strictly positive and* σ_f *is weak-mixing then* $\pi(x) \sim \frac{x}{\log x}$ *as* $x \to \infty$, *where* $\pi(x) = \sum_{N(\tau)\leq x} 1$.

Equivalently, $\#\{\tau : \lambda(\tau) \leq x\} \sim \frac{e^{hx}}{hx}$.

The asymptotic formula $\pi(x) \sim \frac{x}{\log x}$ gives some additional information on the distribution of $\lambda(\tau)$ in remote intervals. Specifically, we can use Stieltjes integration with respect to $\pi(x)$ to get

$\sum_{N(\tau)\leq x} e^{iah\lambda(\tau)} = \int_1^x y^{ia} d\pi(y)$, and after some easy manipulations one obtains

$$\sum_{N(\tau)\leq x} e^{iah\lambda(\tau)} \sim \frac{x^{ia}}{(1+ia)} \pi(x) .$$

Hence

$$\sum_{x<\lambda(\tau)\leq x+1} e^{iah\lambda(\tau)} \sim e^{iahx} \frac{e^{hx}}{hx} \left[\frac{e^{h(1+ia)} - 1}{1+ia} \right]$$

and

$$\sum_{x<\lambda(\tau)\leq x+1} 1 \sim \frac{e^{hx}}{hx} (e^h - 1) .$$

Thus

$$\frac{\sum_{x<\lambda(\tau)\leq x+1} e^{iah\lambda(\tau)}}{\sum_{x<\lambda(\tau)\leq x+1} 1} \sim \frac{e^{iahx}\,(e^{h(1+ia)}-1)}{(e^{h}-1)(1+ia)} .$$

If we choose $ah = 2\pi k$, $k \in \mathbb{Z}$ then we get

$$\frac{\sum_{x<\lambda(\tau)\leq x+1} e^{2\pi ik\lambda(\tau)}}{\sum_{x<\lambda(\tau)\leq x+1} 1} \sim e^{2\pi ikx} \frac{h}{h+2\pi ik}$$

and the latter is the Fourier transform of the probability density $\dfrac{he^{hy}}{(e^{h}-1)}$ translated through an angle $2\pi x$.

In particular, if we let $x \to +\infty$ (through $\mathbb{Z}$) then we see that $\{\lambda(\tau) : x < \lambda(\tau) \leq x+1\}$ is distributed as the probability density $\dfrac{he^{hy}}{(e^{h}-1)}$.

Notes

The suspended flow construction is classical in ergodic theory. Suspended flows over subshifts of finite type occur throughout the symbolic dynamics of Bowen [15] and Ratner [73].

Abramov's work is contained in two articles published in 1961 [1], [2].

Proposition 6.1 occurs in the article of Bowen and Ruelle [17].

The equivalence of (i) and (ii) in Proposition 6.2 was shown by Bowen in [15], by a somewhat different method.

Theorem 6.3, and Corollary 6.3.1, occur in the article of Parry and Pollicott [66]. The first part of Theorem 6.4 occurs in an article of Pollicott [71], whereas the second part occurs in an article of Parry [67] (as does Corollary 6.4.1).

The asymptotic analysis of $\psi(x)$ and $\pi'(x)$ are taken from the article of Parry [65], dealing with locally constant suspended flows - but the method is standard in number theory.

Theorem 6.9 and the preceding lemmas are taken from the article of Parry and Pollicott. The proof is modelled on the classical proof of the prime number theorem, for which good references are [30], [54]. The Ikehara-Wiener theorem is proved in [30], [103], for example - and a proof is presented in Appendix I, for the convenience of the reader.

CHAPTER 7

EQUIDISTRIBUTION THEOREMS FOR SUSPENDED FLOWS

In the previous chapter we showed that there exists a very simple asymptotic formula for the number of closed orbits of a weak mixing suspended flow. Here we show that these closed orbits exhibit a regularity in a spatial sense. In particular, we show them to be equidistributed relative to the measure of maximal entropy, in a very natural way. We shall also prove weighted versions of these results corresponding to more general equilibrium states.

We continue with the assumption that σ_f is weak mixing where $f \in F_\theta$ is strictly positive. Let $G, K \in C(X_f)$ be such that

$$g(x) = \int_0^{f(x)} G(x,t)dt \in F_\theta \qquad k(x) = \int_0^{f(x)} K(x,t)dt \in F_\theta .$$

In Chapter 6 we introduced the zeta function in two variables

$$\zeta(s,z) = \exp \sum_{n=1}^{\infty} \frac{1}{n} \sum_{\mathrm{Fix}_n} e^{g^n - csf^n + zk^n}$$

where $c = P(G) > 0$, and showed in Corollary 6.4.1 that

$$\eta(s) = \frac{\zeta_2(s,0)}{\zeta(s,0)} = \frac{\int K dm_f}{c(s-1)} + \alpha(s)$$

where $\alpha(s)$ is analytic in $\mathcal{R}(s) \geq 1$ and m_f is the equilibrium state of G.

The demonstration that $\zeta_{-hf}(s)$ has an Euler product presentation (in Chapter 6) can be easily modified to show a similar result for $\zeta(s,z)$. In particular, if $\mathcal{R}(s) > 1$ and $|z|$ is sufficiently small (depending on s) then

$$\zeta(s,z) = \prod_{\tau}(1 - e^{(\lambda_G(\tau)-cs\lambda(\tau)+z\lambda_K(\tau))})^{-1}$$

where $\lambda_G(\tau) = \int_{\tau} G$ and $\lambda_K(\tau) = \int_{\tau} K$ denote the integrals around the closed orbit τ relative to Lebesgue measure. Hence,

$$\begin{aligned}\zeta(s,z) &= \exp - \sum_{\tau} \log (1 - e^{(\lambda_G(\tau)-cs\lambda(\tau)+z\lambda_K(\tau))})^{-1} \\ &= \exp \sum_{n=1}^{\infty} \sum_{\tau} \frac{1}{n} e^{(\lambda_G(\tau)-cs\lambda(\tau)+z\lambda_K(\tau))n} .\end{aligned}$$

Logarithmically differentiating with respect to the z coordinate at $z = 0$ we get

$$\eta(s) = \sum_{n=1}^{\infty} \sum_{\tau} \lambda_K(\tau)\, e^{(\lambda_G(\tau) - cs\lambda(\tau))n}$$

$$= \frac{\int K dm_f}{c(s-1)} + \alpha(s)\,.$$

Our first step is to simplify this expression by showing that

$$\sum_{n=2}^{\infty} \sum_{\tau} \lambda_K(\tau)\, e^{(\lambda_G(\tau) - cs\lambda(\tau))n}$$

is analytic in $\mathcal{R}(s) > 1 - \varepsilon$, for some $\varepsilon > 0$.

Since $c = P(G)$ we know that $P(g-cf) = 0$ by Proposition 6.1. It is easy to see that $g-cf \sim -g'$ where g' is a strictly positive function, say $g' \geq 3c\varepsilon\, \|f\|_\infty > 0$. By Proposition 1.2 $g-cf$ is cohomologous to a function of future coordinates and by the proof of Theorem 2.2 this function is cohomologous to one for which the corresponding Ruelle operator is normalised which suffices to prove the function is strictly negative. We can therefore assume that $\lambda_G(\tau) - c\lambda(\tau) \leq -3\varepsilon c\lambda(\tau) < 0$.

Furthermore, we can assume for convenience that $K \geq 0$, and with $\mathcal{R}(s) = u > 1-\varepsilon$ we can estimate,

$$\sum_{n=2}^{\infty} \lambda_K(\tau)\, e^{(\lambda_G(\tau) - cu\lambda(\tau))n} \leq \lambda_K(\tau)\, \frac{e^{2(\lambda_G(\tau) - cu\lambda(\tau))}}{1 - e^{\lambda_G(\tau) - cu\lambda(z)}}$$

$$\leq C_1 \lambda_K(\tau)\, e^{\lambda_G(\tau) - c\lambda(\tau)}\, e^{\lambda_G(\tau) - c\lambda(\tau)\,(1-2\varepsilon)} \leq C_1 \lambda_K(\tau)\, e^{-3c\lambda(\tau)\varepsilon}\, e^{\lambda_G(\tau) - c\lambda(\tau)\,(1-2\varepsilon)}$$

$$= C_1 \lambda_K(\tau)\, e^{\lambda_G(\tau) - c\lambda(\tau)(1+\varepsilon)}$$

(for some constant $C_1 > 0$, provided $u > 1-\varepsilon$) .

Thus

$$\sum_{n=2}^{\infty} \sum_{\tau} \lambda_K(\tau)\, e^{(\lambda_G(\tau) - cu\lambda(\tau))n} \leq C_1 \sum_{\tau} \lambda_K(\tau)\, e^{\lambda_G(\tau) - c\lambda(\tau)(1+\varepsilon)}$$

$$\leq C_1 \sum_{m=1}^{\infty} \sum_{\tau} \lambda_K(\tau)\, e^{[\lambda_G(\tau) - c\lambda(\tau)(1+\varepsilon)]\, m} = C_1 \eta\, (1+\varepsilon) < \infty\, .$$

Therefore we have the following (without necessarily assuming $K \geq 0$).

PROPOSITION 7.1. $\eta_1(s) = \sum_{\tau} \lambda_K(\tau)\, e^{\lambda_G(\tau) - cs\lambda(\tau)} = \dfrac{\int K dm_f}{c(s-1)} + \alpha_1(s)$ *with* $\alpha_1(s)$ *analytic in* $\mathcal{R}(s) \geq 1$.

Again assume (for convenience) that $K \geq 0$, then defining

$S(x) = \sum_{e^{c\lambda(\tau)} \leq x} \lambda_K(\tau) e^{\lambda_G(\tau)}$ we have

$$\eta_1(s) = \int_1^{\infty} x^{-s}\, dS(x)\, .$$

We can apply the Ikehara Wiener Tauberian theorem to the above proposition and deduce that

$$S(x) \sim x.\frac{\int K dm_f}{c}, \text{ as } x \to \infty. \tag{7.1}$$

Taking the ratio of this quantity with the same expression with $K \equiv 1$ gives the following estimates (without necessarily requiring $K \geq 0$).

PROPOSITION 7.2. $$\frac{\sum_{e^{c\lambda(\tau)} \leq x} \lambda_K(\tau) e^{\lambda_G(\tau)}}{\sum_{e^{c\lambda(\tau)} \leq x} \lambda(\tau) e^{\lambda_G(\tau)}} \to \int K dm_f, \text{ as } x \to \infty$$

or equivalently,

$$\frac{\sum_{\lambda(\tau) \leq x} \lambda_K(\tau) e^{\lambda_G(\tau)}}{\sum_{\lambda(\tau) \leq x} \lambda(\tau) e^{\lambda_G(\tau)}} \to \int K dm_f, \text{ as } x \to \infty.$$

This is a weighted uniform distribution result, where the weights $e^{\lambda_G(\tau)}$ determine the limiting measure, the equilibrium state of G.

From (7.1) we can also deduce that

$$\sum_{\lambda(\tau) \leq x} \lambda_K(\tau)\, e^{\lambda_G(\tau)} \sim \frac{e^{cx}}{c} \int K dm_f$$

from which we obtain

$$\sum_{x<\lambda(\tau)\le x+\varepsilon} \lambda_K(\tau)\, e^{\lambda_G(\tau)} \sim \frac{e^{cx}}{c}\,(e^{c\varepsilon}-1)\int K\,dm_f$$

and therefore we can again divide by the corresponding quantities for $K \equiv 1$, to deduce the following:

PROPOSITION 7.3. $$\frac{\sum_{x<\lambda(\tau)\le x+\varepsilon} \lambda_K(\tau)e^{\lambda_G(\tau)}}{\sum_{x<\lambda(\tau)<x+\varepsilon} \lambda(\tau)e^{\lambda_G(\tau)}} \to \int K\,dm_f \text{ as } x\to\infty .$$

When $G \equiv 0$ *then* m_f *is the measure of maximal entropy and we have an unweighted uniform distribution theorem.*

PROPOSITION 7.4. *If* m_f *is the measure of maximal entropy then*

$$\sum_{\lambda(\tau)\le x} \lambda_K(\tau) \sim \frac{e^{hx}}{h}\int K\,dm_f, \text{ as } x\to\infty$$

and

$$\frac{\sum_{x<\lambda(\tau)\leq x+\varepsilon} \lambda_K(\tau)}{\sum_{x<\lambda(\tau)<x+\varepsilon} \lambda(\tau)} \to \int K dm_f, \text{ as } x \to \infty$$

where h = P(0) *is the topological entropy of the flow.*

We can give a slightly different formulation of these equidistribution results. Clearly, we can write

$$\sum_{\lambda(\tau)\leq x} \frac{\lambda_K(\tau)}{\lambda(\tau)} e^{\lambda_G(\tau)} \geq \frac{1}{x} \sum_{\lambda(\tau)\leq x} \lambda_K(\tau)\, e^{\lambda_G(\tau)}.$$

If we choose any $\gamma > 1$ and set $y = x/\gamma$ then as with the estimates for Theorem 6.9 we have

$$\sum_{\lambda(\tau)\leq x} \frac{\lambda_K(\tau)}{\lambda(\tau)} e^{\lambda_G(\tau)} \geq \sum_{\lambda(\tau)\leq y} \frac{\lambda_K(\tau)}{\lambda(\tau)} e^{\lambda_G(\tau)} + \frac{1}{y} \sum_{y<\lambda(\tau)\leq x} \lambda_K(\tau)\, e^{\lambda_G(\tau)}/e^{c\gamma y}.$$

If we choose $1 < \gamma' < \gamma$, then we estimate

$$\overline{\lim_x} \frac{cx}{e^{cx}} \sum_{\lambda(\tau)\leq x} \frac{\lambda_K(\tau)}{\lambda(\tau)} e^{\lambda_G(\tau)} \leq \gamma\, \overline{\lim_y} \left(\frac{y}{e^{c(\gamma-\gamma')y}} \right) \sum_{\lambda(\tau)\leq y} \lambda_K(\tau)\, e^{\lambda_G(\tau)}/e^{c\gamma' y}$$

$$+\gamma\, \overline{\lim_x} \frac{cx}{e^{cx}} \sum_{n\lambda(\tau)\leq x} \lambda_K(\tau)$$

$$= 0 + \gamma \int K dm_f .$$

$\left(\sum_{\lambda(\tau)\le y} \lambda_K(\tau)\, e^{\lambda_G(\tau)} / e^{\gamma y}\right.$ is bounded for all y, by estimates on the zeta function.) In particular, we arrive at the following equidistribution results:

$$\frac{\sum_{x<\lambda(\tau)\le x+\varepsilon} (\lambda_K(\tau)/\lambda(\tau)) e^{\lambda_G(\tau)}}{\sum_{x<\lambda(\tau)\le x+\varepsilon} e^{\lambda_G(\tau)}} \to \int K dm_f .$$

In the special case $G \equiv 0$ then m_f is the measure of maximal entropy and

$$\frac{\sum_{x<\lambda(\tau)\le x+\varepsilon} \lambda_K(\tau)/\lambda(\tau)}{\sum_{x<\lambda(\tau)\le x+\varepsilon} 1} \to \int K dm_f .$$

Throughout this chapter we have assumed $P(G) > 0$. However, as we shall see in Chapter 11, the case $P(G) = 0$ is important. For this case we write $G_\delta = G + \delta$ where $\delta > 0$ is constant so that $P(G_\delta) = \delta > 0$. Applying Proposition 7.3 to G_δ we obtain

$$\frac{\sum_{x < \lambda(\tau) \leq x+\varepsilon} \lambda_K(\tau) e^{\lambda_G(\tau)} e^{\delta\lambda(\tau)}}{\sum_{x < \lambda(\tau) \leq x+\varepsilon} \lambda(\tau) e^{\lambda_G(\tau)} e^{\delta\lambda(\tau)}} \to \int K dm_f$$

where m_f is the equilibrium state of G (which, of course, is the equilibrium state of G_δ). We get upper and lower estimates for the above ratio by replacing the exponent $\lambda(\tau)$ by x and $x + \varepsilon$. For example, the ratio is bounded above by

$$e^{\delta\varepsilon} \cdot \frac{\sum_{x < \lambda(\tau) \leq x+\varepsilon} \lambda_K(\tau) e^{\lambda_G(\tau)}}{\sum_{x < \lambda(\tau) \leq x+\varepsilon} \lambda(\tau) e^{\lambda_G(\tau)}} .$$

Since $\delta > 0$ is arbitrary we see that Proposition 7.3 holds when $P(G) \geq 0$.

Notes

Spatial equidistribution results for closed orbits of suspended flows were originally due to Bowen [14], [15]. Bowen proved these for hyperbolic flows (without actual asymptotic estimates) and thus for suspended flows (by using the embedding result in [11], say). The approach we take is based on [67].

CHAPTER 8

GALOIS EXTENSIONS AND CHEBOTAREV TYPE THEOREMS

In Chapter 6 we presented an asymptotic formula for closed orbits of a suspended flow which has certain similarities with the prime number theorem. Another classical result from number theory is the Chebotarev theorem; this theorem gives asymptotic formulae for the way in which primes in a given number field split in a finite extension field. We shall consider an analogous situation for hyperbolic flows where instead of field extensions we consider covering or extension spaces. Our aim is to study the distribution of the lifts of closed orbits in terms of the associated Galois group.

Following our usual notation, (X,σ) will be a subshift of finite type. With G a compact Lie group we wish to define a G-extension of σ. Since G has a faithful representation in the group $U(d)$ of unitary $d \times d$ matrices (for some d) we may suppose that G is a closed subgroup of $U(d)$. For a continuous function $\alpha : X \to G$ we define

$$\mathrm{var}_n\, \alpha = \sup \{ |\alpha(x) - \alpha(y)| \; | x_i = y_i , |i| < n \}$$

where $|\ |$ denotes the Euclidean norm of a matrix.

For $0 < \theta < 1$ let $U(d,\theta) = \{\alpha : \alpha$ is continuous and $\mathrm{var}_n\, \alpha \leq K\theta^n$, for $n \in \mathbb{N}$ and some constant $K\}$.

We suppose $\alpha \in U(d,\theta)$ and define $\tilde{\sigma} : \tilde{X} \to \tilde{X}$, where $\tilde{X} = X \times G$, and $\tilde{\sigma}(x,g) = (\sigma x, \alpha(x)g)$. We shall call $\tilde{\sigma}$ a *(Galois) G-extension* of σ. The transformation $\tilde{\sigma}$ commutes with the (free) action of G on $\tilde{X}$, $g : (x,h) \mapsto (x,hg)$; thus we can identify $\sigma = \tilde{\sigma}/G$. We assume, throughout this chapter, that $\tilde{\sigma}$ is topologically transitive.

Given a suspension flow $\sigma_t = \sigma_{f,t}$, $f \in F_\theta$ being strictly positive, we can define in a similar spirit a *(Galois) G-extension* of this flow. Specifically, we define an extension of f to $\tilde{f} : \tilde{X} \to \mathbb{R}$ by $\tilde{f}(x,g) = f(x)$ and let

$$\tilde{X}_f = \{(x,g,t) \in \tilde{X} \times \mathbb{R} : 0 \leq t \leq \tilde{f}(x,g)\},$$

where we identify $(x,g,\tilde{f}(x,g)) \sim (\tilde{\sigma}(x,g),\ 0)$. We define the G-extension flow locally by $\tilde{\sigma}_{f,t}(x,g,u) = (x,g,u+t)$ and extend it using the above identifications.

As in the discrete case G acts on $\tilde{X}_f$ by $g : (x,h,t) \to (x,hg,t)$ and since this action commutes with $\tilde{\sigma}_{f,t}$ we can identify $\tilde{\sigma}_{f,t}/G = \sigma_{f,t}$. In the special case where G is finite $\tilde{\sigma} : \tilde{X} \to \tilde{X}$ is a subshift of finite type, and $\tilde{\sigma}_{f,t}$ is a suspension flow.

We write $\pi : \tilde{X}_f \to X_f$, where $\pi(x,g,t) = (x,t)$. Given a closed orbit τ for $\sigma_{f,t}$ of least period $\lambda(\tau)$ we observe that for $\tilde{p} \in \tilde{X}_f$ with $\pi(\tilde{p}) = p \in \tau$ we have $\pi(\tilde{\sigma}_{f,\lambda(\tau)}\tilde{p}) = p$. This follows from the simple identity $\sigma_{f,t}\pi = \pi\tilde{\sigma}_{f,t}$.

In particular, there exists a unique element $g \in G$ such that $g\tilde{p} = \tilde{\sigma}_{f,\lambda\tau)}(\tilde{p})$. If we choose another point in $\tilde{X}_f$ which projects to $p \in \tau \subseteq X_f$ then it must be of the form $h\tilde{p}$, for some $h \in G$. Since the action of G commutes with the flow we have $\tilde{\sigma}_{f,\lambda(\tau)}h\tilde{p} = hg\tilde{p} = (hgh^{-1})h\tilde{p}$. Thus the action of hgh^{-1} takes the lift $h\tilde{p}$ of p to $\tilde{\sigma}_{f,\lambda(\tau)}h\tilde{p}$.

Hence to any closed σ_f-orbit τ in X_f we can associate a well-defined conjugacy class $[g]$ in G, called the *Frobenius class* of τ. We shall denote the Frobenius class of τ by $[\tau]$.

In this chapter we shall be interested in how closed orbits are distributed according to their Frobenius classes. This involves some modifications to the analysis of the previous chapters. Whereas the proof of the prime orbit theorem uses zeta functions modelled on the Riemann zeta function we shall introduce for this type of analysis analogues of Artin's L-functions.

Let $R_\chi : G \to U(d)$ be a finite dimensional unitary representation of the compact Lie group G, with character $\chi : G \to \mathbb{C}$ (i.e. $\chi = \text{trace}\, R_\chi$). We define the *L-function* of χ by

$$L(s,\chi) = \prod_\tau \det\left(I - \frac{R_\chi[\tau]}{N(\tau)^s}\right)^{-1} = \exp \sum_{n=1}^{\infty} \frac{1}{n} \sum_\tau \frac{\chi([\tau]^n)}{N(\tau)^{sn}},$$

which can be seen to converge for $\mathcal{R}(s) > 1$ by comparison with the zeta function $\zeta(s)$.

The L-function only depends on the conjugacy class of R_χ which is determined by the character χ. Thus although $[\tau]^n$ only determines an element of G up to conjugacy it uniquely determines the value $\chi[\tau]^n$. Functions on G which are constant on conjugacy classes, such as characters, are usually called *class functions.*

Given two characters χ_1, χ_2 we note that

$$\begin{aligned}\log L(s, \chi_1 + \chi_2) &= \sum_{n=1}^{\infty} \frac{1}{n} \sum_{\tau} \frac{\chi_1([\tau]^n) + \chi_2([\tau]^n)}{N(\tau)^{sn}} \\ &= \log L(s, \chi_1) + \log L(s, \chi_2).\end{aligned}$$

We therefore have the following:

PROPOSITION 8.1. *For characters* χ_1, χ_2 *we have*

$$L(s, \chi_1 + \chi_2) = L(s, \chi_1)\ L(s, \chi_2).$$

Furthermore, $L(s, \chi_0) = \zeta(s)$ *where* χ_0 *is the trivial (principal) character corresponding to the one-dimensional representation* $g \to 1$ *for all* $g \in G$. *By analogy with the zeta function we can use the correspondence between* σ-periodic *orbits* $\{x, \sigma x, ..., \sigma^{n-1} x\}$ *and* σ_f-periodic *orbits* τ *of least period* $\lambda(\tau) = f^n(x)$ *to write:*

$$L(s,\chi) = \exp \sum_{n=1}^{\infty} \frac{1}{n} \sum_{Fix_n} \chi(\alpha_n(x))\, e^{-shf^n(x)} \tag{8.1}$$

where $\alpha_n(x) = \alpha(\sigma^{n-1}x)...\alpha(\sigma x)\,\alpha(x)$.

To analyse the domain of (8.1) it is convenient to move to the setting of one-sided shifts. As we have seen in Proposition 1.2 it is possible to assume, without loss of generality, that f is a function depending only on future coordinates. There is a similar result we can apply to the function $\alpha : X \to G$ to show that, without loss of generality, we may assume that α depends only on future coordinates.

PROPOSITION 8.2. *If* $\alpha \in U(d,\theta)$ *then there exists* $\gamma \in U(d,\theta^{\frac{1}{2}})$ *such that:*

$$\alpha' = (\gamma \circ \sigma)^{-1}\, \alpha\gamma \in U(d,\theta^{\frac{1}{2}})$$

depends only on future coordinates (i.e. $\alpha'(x) = \alpha'(y)$ *if* $x_n = y_n$ *for* $n \geq 0$). *Hence, for any character* χ, *and* $\sigma^n x = x$, *we have*

$$\chi(\alpha'_n(x)) = \chi(\gamma(x)^{-1}\, \alpha_n(x)\, \gamma(x)) = \chi(\alpha_n(x)).$$

The proof of this proposition (which is modelled on that of Proposition 1.2) is given in Appendix II.

In particular, we see that we may replace α by α' in (8.1) and assume f,α depend only on future coordinates. The benefit of this is that we can introduce a suitable variant of the Ruelle operator.

Let $F^+(d,\theta)$ be the space of continuous functions $k : X^+ \to \mathbb{C}^d$ for which $\mathrm{var}_n k \leq K\theta^n$, for some $K > 0$, where

$$\mathrm{var}_n k = \sup \{|k(x) - k(y)| : x_i = y_i, 0 \leq i \leq n-1\}.$$

(Here $|\ |$ denotes the Euclidean norm on $\mathbb{C}^d$.)

The space $F^+(d,\theta)$ becomes a Banach space with respect to the norm: $\|k\|_\theta = |k|_\infty + |k|_\theta$ where

$$|k|_\infty = \sup \{|k(x)| : x \in X^+\}, |k|_\theta = \inf \{K : \mathrm{var}_n k \leq K\theta^n\}.$$

For a unitary representation $R : G \to U(d)$ we define

$$(L_{s,R} w)(x) = \sum_{\sigma y = x} e^{-shf(y)} R(\alpha(y)) w(y).$$

As for the case of zeta functions (in Chapter 5) we can construct a meromorphic extension of $L(s,\chi)$ by first studying the spectrum of $L_{s,R}$. By mimicking the proofs of Chapter 4 we first get the following extension of Theorem 4.5.

THEOREM 8.1. *For* $s = u+it$ *and* R *a unitary representation in* $U(d)$ *we have* $\rho(L_{s,R}) \leq e^{P(-uhf)}$.

Since by Proposition 6.1 we know that $P(-hf) = 0$ and $P(-uhf) < 0$ for $u > 1$ we deduce that $\rho(L_{s,R}) \leq 1$ for $u \geq 1$ and $\rho(L_{s,R}) < 1$ for $u > 1$.

As R is unitary we have $\sum_{Fix_n} e^{-shf^n(x)}\chi(\alpha_n(x))$ bounded by $d \sum_{Fix_n} e^{-uhf^n(x)}$ when $s = u + it$. Consequently if $u > 1$ then we have in analogy with Theorem 5.4 that $L(s,\chi)$ converges to a non-zero analytic function. For the case $u = 1$ we have the following version of Theorem 5.6:

THEOREM 8.2. *Assume* $u = 1$ *and* $\rho(L_{s,R}) < 1$. *Then there exists* $\varepsilon > 0$ *such that* $L(s,\chi)$ *is nowhere zero and analytic in* $D_\varepsilon(s) = \{z \in \mathbb{C} : |s-z| < \varepsilon\}$.

It remains to consider the possibility that $\rho(L_{s,R}) = 1$ for $u = 1$. We can modify the proof of Theorem 4.5 to get the following result.

THEOREM 8.3. *If* $u = 1$, $\rho(L_{s,R}) = 1$ *and* $\Re(s) = 1$ *then* $L_{s,R}$ *has a (simple) eigenvalue of modulus one.*

Assume that $s = 1+it_0$, then there exists $w \in F^+(d,\theta)$ such that $L_{(1+it_0),R}w = e^{ib}w$, $b \in \mathbb{R}$, by the above theorem. We can assume for simplicity that $R = R_\chi$, where χ is irreducible.

By the usual convexity-type argument (as in Chapter 4) we conclude that

$$e^{-ht_0 f} R(\alpha)w = e^{ib} w \circ \sigma . \tag{8.2}$$

To see this we must go through the usual argument of assuming $\sum_{\sigma y = x} e^{-hf(y)} = 1$, by changing f by a coboundary, as in the proof of Theorem 2.2, and then

$$|w| = |L_{(1+it_0),R} w| \leq \sum_{\sigma y = x} e^{-hf(y)} |R(\alpha(y)).w(y)| \leq |w| ,$$

which can only be satisfied with the validity of (8.2). Define $F(x,g) = R(g)^{-1}w(x)$ so that $F(\sigma x, \alpha(x)g) = e^{-ht_0 f(x) - ib} F(x,g)$. By comparing the coordinates of this equation and using the topological transitivity of $\tilde{\sigma}$ we see that $F(x,g) = R(g)^{-1}w(x) = \theta(x,g)w_0$ for some constant vector w_0 and continuous function θ. Hence, by fixing $x = x_0$, we deduce that R is one-dimensional since R is irreducible. Thus:

$$e^{-iht_0 f} \chi(\alpha)w = e^{ib} w \circ \sigma, \; w \in F_\theta^+ . \tag{8.3}$$

Summarising we have:

PROPOSITION 8.3. *If* $L(s,\chi)$ *has a pole on* $\mathcal{R}(s) = 1$ *then* R_χ *is one-dimensional (i.e.* $R_\chi = \chi$*).*

Assuming (8.3) is satisfied we have two possible cases:

Case (a): ($e^{ib} \neq 1$). The identity (8.3) is incompatible with the general criteria of zeta functions having poles at $s = 1+it_0$, by Theorem 5.5. In particular, we see that this case is void.

Case (b): ($e^{ib} = 1$). As $e^{-iht_0 f}\chi(\alpha)w = w\circ\sigma$ we define $F : \tilde{X} \to \mathbb{C}$ by $F(x,g) = \chi(g^{-1})w(x)$ to obtain

$$F\tilde{\sigma}(x,g) = e^{-iht_0 f(x)}\, F(x,g)\,. \tag{8.4}$$

Finally, if we define $H : \tilde{X}_f \to \mathbb{C}$ by $H(x,g,t) = e^{-iht_0 t}\, F(x,g)$ then we have $H\tilde{\sigma}_{f,t} = e^{-iht_0 t}\, H$.

If we make the additional assumption that $\tilde{\sigma}_{f,t}$ is topologically weak-mixing (i.e. $H\tilde{\sigma}_{f,t} = e^{ia}H$, has no non-trivial continuous solutions) then this condition is contradictory, and this second case is also void.

THEOREM 8.4. *If* $\tilde{\sigma}_{f,t}$ *is topologically weak-mixing, then for any non-trivial irreducible character* $L(s,\chi)$ *is nowhere vanishing and analytic in* $\mathcal{R}(s) \geq 1$.

When χ is trivial $L(s,\chi)$ reduces to $\zeta(s)$ and we know $\zeta(s)$ has a simple pole at $s = 1$ by Theorem 6.3.

If $\tilde{\sigma}_{f,t}$ is not topologically mixing then the situation is slightly different. We shall postpone a discussion of this case until we have explored the implications of the above theorem for the distribution of closed orbits.

Distribution of Frobenius classes. We can use Theorem 8.4 to prove the following result about the distribution of closed orbits for σ_f according to their Frobenius classes:

THEOREM 8.5. (Chebotarev theorem, weak-mixing case.)

If $\tilde{\sigma}_f$ is topologically weak mixing then for each continuous class function $F \in C(G,\mathbb{C})$ *we have*

$$\sum_{N(\tau)\leq x} F([\tau]) \sim \pi(x)\int F(g)dg \tag{8.5}$$

where dg *denotes the Haar measure on* G. *(We recall that a class function is constant on conjugacy classes in* G *i.e.* $F(ghg^{-1}) = F(h)$ *for all* $h, g \in G$.)

PROOF. The method of proof is similar to that of the prime orbit theorem in Chapter 6. It suffices to prove (8.5) for linear combinations of characters since these span the space of class functions. Without loss of generality we may also assume $F > 0$.

If $\chi \equiv \mathbb{1}$ is the trivial character then $\int \chi dg = 1$ and

$$\frac{L'(s,\chi)}{L(s,\chi)} = \frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty}\sum_{\tau} \frac{\log N(\tau)}{N(\tau)^{sn}} . \tag{8.6}$$

By Theorem 6.3 we see that this is analytic for $\mathcal{R}(s) \geq 1$, except for a simple pole at $s = 1$.

For $\chi \neq \mathbb{1}$ we have $\int \chi dg = 0$ and the logarithmic derivative takes the form:

$$\frac{L'(s,\chi)}{L(s,\chi)} = \sum_{n=1}^{\infty} \sum_{\tau} \chi([\tau]^n) \frac{\log N(\tau)}{N(\tau)^{sn}} . \tag{8.7}$$

Assume that $F = \sum_{i=0}^{N} a_i \chi_i$, with $\chi_0 \equiv \mathbb{1}$. We see that $\int F dg = a_0$ and from (8.6) and (8.7) we get:

$$\sum_{n=1}^{\infty} \sum_{\tau} \frac{\log N(\tau)}{N(\tau)^{ns}} F([\tau]^n) = \frac{a_0}{s-1} + \psi(s) , \tag{8.8}$$

where $\psi(s)$ is analytic in the region $\mathcal{R}(s) \geq 1$.

As in Chapter 6 we may write the lefthand side of (8.8) as $\int t^{-s} d\pi_F(t)$, where $\pi_F(t) = \sum_{N([\tau]^n) \leq t} \log N(\tau) F([\tau]^n)$. By applying the Ikehara-Wiener Tauberian theorem (Theorem 6.7)

$$\pi_F(t) \sim a_0 t = \left(\int F dg\right) t . \tag{8.9}$$

Following the arguments in Chapter 6 we see that (8.9) implies $\sum_{N(\tau)\leq t} F([\tau]) \sim (\int F dg)\pi(t)$.

We return now to the situation where σ_f is weak-mixing but $\tilde{\sigma}_f$ is not weak-mixing. Regarding the expression (8.3) (in case (b)) we cannot automatically discount non-trivial solutions.

In particular, if $\sigma^n x = x$ then $\chi(\alpha_n(x)) = e^{iaf^n(x)}$, where $a = ht_0$. Since we can write $\lambda(\tau) = f^n(x)$ this expression is equal to $\chi([\tau]) = e^{ia\lambda(\tau)}$, for all closed orbits τ of σ_f.

Conversely, if $\chi([\tau]) = e^{ia\lambda(\tau)}$, for all closed orbits τ then we can deduce $w(\sigma x)\, \chi(\alpha(x)) = e^{iaf(x)} w(x)$, for some $w \in F_\theta$, by Proposition 3.7 of Chapter 3.

We shall call a one-dimensional representation $\chi : G \to \mathbb{C}$ *special* if there exists some real number b such that $\chi([\tau]) = e^{ib\lambda(\tau)}$, for all closed orbits τ. The set of such characters is an abelian group. We write χ_b for such a character.

The values b correspond to eigenfrequencies for $\tilde{\sigma}_f$, and so the existence of non-trivial characters χ_b entails $\tilde{\sigma}_f$ being *not* weak-mixing (with $b = ht_0$ in case (b) above).

Since we assume $\tilde{\sigma}_f$ is topologically transitive the map $b \mapsto \chi_b$ is well defined. For otherwise we could find a non-zero character χ with $\chi([\tau]) = 1$, or equivalently $w(\sigma x) . \chi(\alpha(x)) = w(x)$ for some $w \in F_\theta^+$, as explained above. By defining $f(x,g) = \chi(g)w(x)$ we have a $\tilde{\sigma}$-invariant function on $\tilde{X}$ and therefore, this function must be constant, i.e. χ must be trivial, giving a contradiction.

Consider those $b \in \mathbb{R}$ which give rise to the trivial representation $\mathbb{1} : G \to \mathbb{C}$ (i.e. the kernel of $b \mapsto \chi_b$). Any such b will satisfy $e^{ib\lambda(\tau)} = 1$ for all closed orbits τ and as before, we can deduce $w(\sigma x) = e^{ibf(\tau)}w(x)$, for some $w \in F_\theta^+$. If we define $F : X_f \to \mathbb{R}$ by $F(x,t) = e^{ibt}w(x)$ then $F\sigma_{f,t} = e^{ibt}F$, i.e. b is an eigenfrequency for σ_f. In conclusion:

PROPOSITION 8.4. *If $\tilde{\sigma}_f$ is topologically transitive, a necessary and sufficient condition for the existence of non-trivial special characters is that $\tilde{\sigma}_f$ is not weak-mixing. There is a homomorphism from the eigenfrequencies of $\tilde{\sigma}_f$ to the special characters, whose kernel consists of the eigenfrequences of σ_f.*

If $\tilde{\sigma}_f$ is not weak-mixing then we would not expect a simple asymptotic formula like (8.5) to be valid. If χ_b is a (non-trivial) special character we can write

$$\sum_{N(\tau)\le x} \chi_b([\tau]) = \sum_{N(\tau)\le x} e^{ib\lambda(\tau)} .$$

By our estimates in Chapter 6 this expression is asymptotic to $\dfrac{x^{ib}}{(1+ib)}\,\pi(x)$.

When $\tilde{\sigma}_f$ is not weak-mixing $L(s,\chi_b) = \zeta(s-ib/h)$ and since $\zeta(s)$ has a simple pole at $s = 1$ clearly $L(s,\chi_b)$ must have a simple pole at $s = 1 + ib/h$ which explains the non-uniformity of distribution when $\tilde{\sigma}_f$ is not weak mixing.

There is one very simple situation where the assumption that σ_f is weak-mixing forces $\tilde{\sigma}_f$ to be weak-mixing. This is the case where G is a finite group. In particular, we have the following version of Theorem 8.5.

THEOREM 8.6. (Chebotarev density theorem - finite extension case.)

If G *is finite and* $\tilde{\sigma}_f$ *is a topologically transitive* G-extension *of* σ_f *then for each class function* F:

$$\sum_{N(\tau)\leq x} F([\tau]) \sim \pi(x)\int F(g)dg\,.$$

In particular, for each conjugacy class C *of* G

$$\#\,\{\tau : N(\tau) \leq x\,,\,[\tau] = C\} \sim \pi(x)\,\frac{\mathrm{Card}C}{\mathrm{Card}G}\,.$$

(The asymptotic formula for $\pi(x)$ *is given in Chapter 7, and depends on whether* σ_f *is weak-mixing or not.)*

We have deduced Theorem 8.6 from the general case of Theorem 8.5.

However, it is possible to give a direct proof in keeping with the original Chebotarev theorem (for finite extensions) in number theory. We shall briefly indicate the main ideas.

Since G is finite so is the set of irreducible characters. If we define

$$\zeta_C(s) = \prod_{\substack{\tau \\ [\tau]=C}} (1-N(\tau)^{-s})^{-1}$$

then with

$$L(s,\chi) = \prod_{\tau} \det\left(I-R_\chi([\tau])\, e^{-sh\lambda(\tau)}\right)$$

we have:

$$\zeta'_C(s)/\zeta_C(s) = \frac{|C|}{|G|} \sum_{\chi} \chi(g^{-1}) \frac{L'(s,\chi)}{L(s,\chi)}\ , \text{ for } g \in C\,.$$

(By expanding the logarithmic derivatives and using the orthogonality of the characters.)

When $\chi \neq \mathbb{1}$ the analogue of Theorem 8.4 applies and we have $\zeta'_C(s)/\zeta_C(s)$ is analytic for $\mathcal{R}(s) \geq 1$, except for a simple pole at $s = 1$ with residue $\frac{|C|}{|G|}$ (coming from the zeta function $\zeta(s) = L(s,\mathbb{1})$ when $\chi \equiv \mathbb{1}$). Theorem 8.6 then follows closely the derivation of the original prime orbit theorem in Chapter 6.

Notes

The results in this chapter are based on the article of Parry-Pollicott [68]. Earlier results for constant curvature geodesic flows are due to Sunada [98] and Sarnak [84], and for variable curvature geodesic flows the results were proved by Adachi-Sunada [3].

The motivation for these extension results is the Chebotarev density theorem in number theory (cf. Cassels-Fröhlich [22]).

The case of extensions by compact Lie groups, which forms the bulk of the chapter, was dealt with in sections 9 and 10 of the article of Parry-Pollicott [68].

The finite extension case, which we briefly describe at the end of the chapter as a corollary to the compact case, is analysed in detail in [68].

CHAPTER 9

APPLICATIONS TO HYPERBOLIC FLOWS

To date we have concentrated on analysing zeta functions and proving distribution results for closed orbits in the context of suspended flows. In this chapter we shall show how these ideas and results can be transferred to the "more natural" setting of hyperbolic flows. These include as special cases Axiom A flows (as studied by Smale) and Anosov flows (as studied by the Russian school), and in particular the canonical example in ergodic theory - the geodesic flow on the unit tangent bundle of a compact manifold with negative sectional curvatures.

The transition from the theory of suspended flows to that of hyperbolic flows follows standard lines based on ideas of Bowen, Ratner, Ruelle, Sinai and others. The idea is to introduce Poincaré sections for the flow with an additional "Markov" property. This enables the Poincaré map between the sections to be closely modelled by a subshift of finite type. The return time between the Poincaré sections then corresponds to a roof function. In this way we can model the hyperbolic flow by a suspended flow, of the type we have already described. For the convenience of the reader we have summarised this standard, but somewhat complicated construction, in Appendix III.

In studying properties of a hyperbolic flow the basic procedure we follow is to establish the corresponding result for an associated suspended flow, and then to transfer the result back to the hyperbolic flow. In the preceding chapters we have established a number of results on zeta functions and closed orbits for hyperbolic flows. In this chapter we shall establish the corresponding results for hyperbolic flows.

We begin by recalling the definition of a *hyperbolic flow* $\varphi_t : \Lambda \to \Lambda$ on a *basic set* Λ:

Let $\varphi_t : M \to M$ be a C^1 flow on a compact C^∞ manifold and let $\Lambda \subset M$ be a φ-invariant compact set such that

(i) There exists a splitting $T_1 M = E^0 \oplus E^s \oplus E^u$ such that

(a) there exist constants $C, \lambda > 0$ with $\|D\varphi_t|_{E^s}\|, \|D\varphi_{-t}|_{E^u}\| \leq Ce^{-\lambda t}, t \geq 0$

(b) E^0 is one-dimensional and tangent to the flow.

(ii) Λ contains a dense orbit.

(iii) The periodic orbits in Λ are dense (and Λ consists of more than a single closed orbit).

(iv) There exists an open set $U \supset \Lambda$ such that $\Lambda = \bigcap_{t=-\infty}^{\infty} \varphi_t U$.

We observe the connection with Smale's work on Axiom A flows: A C^1 flow $\varphi_t : M \to M$ on a compact manifold satisfies *Axiom A* if the non-wandering set $\Omega = \{x \in M :$ for each neighbourhood $U \ni x$, there exists $t_n \nearrow +\infty$ with $\varphi_{t_n} U \cap U \neq 0\}$ satisfies (i) and (iii) above. In particular, Smale showed that for an Axiom A flow Ω is a finite union of basic sets, hyperbolic closed orbits and hyperbolic fixed points.

PROPOSITION 9.1. $\varphi_t : \Lambda \to \Lambda$ *has a unique measure of maximal entropy which we denote by* m.

(The proof of this result is given in Appendix III.)

We say that the hyperbolic flow of $\varphi_t : \Lambda \to \Lambda$ is (*topologically*) *weak-mixing* if there is no non-trivial solution to $F\varphi_t = e^{iat}F$, $a \in \mathbb{R}$, $F \in C(\Lambda)$. We denote the topological entropy of $\varphi_t : \Lambda \to \Lambda$ by $h(\varphi) = \sup \{h_\mu(\varphi_1) : \mu$ is a φ-invariant probability measure$\}$.

We summarise below how a suspended flow (of finite type) associated to a hyperbolic flow via the Markov sections closely models the original flow. For the reader's convenience we have sketched the proofs of these results in Appendix III, where references can also be found for complete proofs.

LEMMA 9.1. *We can associate to a hyperbolic flow* $\varphi_t : \Lambda \to \Lambda$ *a suspended flow* $\sigma_{f,t} : X_f \to X_f$ *and a continuous map* $\pi : X_f \to \Lambda$ *such that*

(i) π *is surjective and* $\varphi_t\pi = \pi\sigma_{f,t}$.

(ii) π *is bounded-one, and one-one on a residual set.*

(iii) π *is an isomorphism (with respect to the unique measures of maximal entropy) and in particular, they have the same topological entropy* $h(\sigma_f) = h(\varphi) = h$.

(iv) φ *is topologically weak-mixing if and only if* σ_f *is topologically weak-mixing.*

At the measurable level $\sigma_{f,t} : X_f \to X_f$ is a perfect model for $\varphi_t : \Lambda \to \Lambda$, namely an isomorphic representation and at the topological level φ_t and $\sigma_{f,t}$ are as close to being conjugate as one can reasonably expect. (In general, we cannot expect π to be a homeomorphism since X_f is always one-dimensional, whereas Λ need not be.)

We want to present the main results of these notes for hyperbolic flows and we begin with the results for zeta-functions.

Let τ denote a closed φ-orbit of least period $\lambda(\tau)$. There is a denumerable infinity of closed orbits and we define the *zeta-function* to be the function of a complex variable $s \in \mathbb{C}$ given by the Euler product:

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

(We shall show below that this converges to a non-zero analytic function for $\mathcal{R}(s) > h(\varphi)$.)

Of course, the map π does not, in general, give a one-one correspondence between closed orbits for φ and σ_f and so an identification of the zeta function $\zeta_\varphi(s)$ and $\zeta_{\sigma_f}(s)$ is generally impossible. However, there is a very explicit relationship of the following form:

LEMMA 9.2. *There exists a finite family of suspended flows* σ_{f_i} $(i=1,\ldots,N)$ *with* $h(\sigma_{f_i}) < h(\sigma_f) = h(\varphi)$ *such that*

$$\zeta_\varphi(s) = \zeta_{\sigma_f}(s) \frac{\prod_{i=1}^{p} \zeta_{\sigma_{f_i}}(s)}{\prod_{i=p+1}^{N} \zeta_{\sigma_{f_i}}(s)} \tag{9.1}$$

(The proof is outlined in Appendix III.)

By the results of Chapter 6 we see that $\zeta_{\sigma_{f_i}}(s)$ is non-zero and analytic for $\mathcal{R}(s) > h(\varphi) - \varepsilon = \max_i \{h(\sigma_{f_i})\}$. (This also proves our claim that $\zeta_\varphi(s)$ is non-zero and analytic for $\mathcal{R}(s) > h(\varphi)$.) In particular, Lemma 9.2 and Lemma 9.1 (iii), (iv) give the following analogue of Theorem 6.3.

THEOREM 9.1.

If φ *is topologically weak-mixing then* $\zeta_\varphi(s)$ *has a non-zero analytic extension to a neighbourhood of* $\mathcal{R}(s) \geq h(\varphi)$, *except for a simple pole at* $s = h$.

There is a weighted version of the zeta-function, which we define as follows. Let $F : \Lambda \to \mathbb{R}$ be a C^α function. We define

$$\zeta_\varphi(s,F) = \prod_\tau (1-e^{\lambda_F(\tau)-s\lambda(\tau)})^{-1}$$

where $\lambda_F(\tau) = \int_0^{\lambda(\tau)} F(\varphi_t x)dt$, for any point $x \in \tau$.

We define the *pressure* $P(F) = \sup \{h_\mu(\varphi_1) + \int F d\mu : \mu$ is φ-invariant probability measure$\}$.

The natural generalisation of Lemma 9.2 is also true:

LEMMA 9.3. *There exists a finite family of suspended flows* σ_{f_i} *and functions* $F_i = F \circ \pi_i$ *(where* $\pi_i : X_{f_i} \to \Lambda$ *is as defined in Appendix III) with* $P(F_i) < P(F)$ *and*

$$\zeta_\varphi(s,F) = \zeta_{\sigma_f}(s,F \circ \pi) \frac{\prod_{i=1}^{P} \zeta_{\sigma_{f_i}}(s,F_i)}{\prod_{i=P+1}^{N} \zeta_{\sigma_{f_i}}(s,F_i)} . \tag{9.2}$$

This leads to the following extension of theorem 9.1.

THEOREM 9.2. $\zeta_\varphi(s,F)$ *is non-zero and analytic for* $\mathcal{R}(s) \geq P(F)$ *except for a simple pole at* $s = P(F)$.

We turn now to the three distribution results for closed orbits. (For the suspended flow these are contained in Chapters 6, 7 and 8.) There are two alternative methods for establishing these results for hyperbolic flows. The first approach is to repeat, almost verbatim, the proofs in Chapters 6, 7 and 8 for suspended flows after replacing the necessary conditions on the zeta-functions for suspended flows by the corresponding results for the hyperbolic flow in Theorems 9.1 and 9.2. The second approach is to show that counting functions for the hyperbolic and associated suspended flow are asymptotic. This is true since the (negligible) difference is due to closed φ orbits which pass through the boundaries of Markov sections.

Our first result is the analogue of Theorems 6.9 and 6.5 for hyperbolic flows. Let $\pi'(x)$ be the number of closed orbits of least period $\lambda(\tau) \leq x$, $x > 0$.

THEOREM 9.3. (Prime orbit theorem for hyperbolic flows.)

(a) *If* $\varphi_t : \Lambda \to \Lambda$ *is a topologically weak-mixing hyperbolic flow then*

$$\pi'(t) \sim \frac{e^{hx}}{hx} .$$

(b) *If* $\varphi_t : \Lambda \to \Lambda$ *is not weak-mixing then*

$$\pi'(t) \sim \frac{e^{h}}{e^{h} - 1} \frac{e^{h[t]}}{[t]} .$$

(where [t] *denotes the integer part of* t).

We consider next the appropriate analogue of the equidistribution theorems.

In view of the above result and Lemma 9.1 (iii) we have the following version for hyperbolic flows.

THEOREM 9.4. (Equidistribution theorem for hyperbolic flows.)

Let $\varphi_t : \Lambda \to \Lambda$ *be a hyperbolic flow and let* $G : \Lambda \to \mathbb{R}$ *be a Hölder continuous function with equilibrium state* m. *If* $F \in C(\Lambda)$ *then*

$$\frac{\sum_{\lambda(\tau)\le t} \lambda_F(\tau) . e^{\lambda_G(\tau)}}{\sum_{\lambda(\tau)\le t} \lambda(\tau)\, e^{\lambda_G(\tau)}} \to \int F dm \text{ as } t \to \infty \text{; and}$$

$$\frac{\sum_{\lambda(\tau)\le t} \frac{\lambda_F(\tau)}{\lambda(\tau)} . e^{\lambda_G(\tau)}}{\sum_{\lambda(\tau)\le t} e^{\lambda_G(\tau)}} \to \int F dm, \text{ as } t \mapsto \infty .$$

We observe that if μ is the unique equilibrium state for the suspended flow σ_f and $G \circ \pi : X_f \to \mathbb{R}$ then by Lemma 9.1 (iii) we have $\int F dm = \int (F \circ \pi) d\mu$, which makes the deduction from Chapter 7 straightforward.

COROLLARY 9.4.1. *Let* $\varphi_t : \Lambda \to \Lambda$ *be a hyperbolic flow with unique measure of maximal entropy* m. *If* $F \in C(\Lambda)$ *then:*

$$\frac{\sum_{\lambda(\tau)\le t} \lambda_F(\tau)}{\sum_{\lambda(\tau)\le t} \lambda(\tau)} \to \int F dm, \text{ as } t \to \infty \text{; and}$$

$$\frac{1}{\pi'(t)} \sum_{\lambda(\tau)\le t} \frac{\lambda_F(\tau)}{\lambda(\tau)} \to \int F dm, \text{ as } t \to \infty .$$

Finally, we want to give the Chebotarov density theorem for hyperbolic flows.

Let G be a compact Lie group and let $\tilde{\varphi}_t : \tilde{\Lambda} \to \tilde{\Lambda}$ be a (Galois) G-extension of the hyperbolic flow $\varphi_t : \Lambda \to \Lambda$ with projection $\pi : \tilde{\Lambda} \to \Lambda$. Given a closed orbit $\tau \subset \Lambda$ and $p \in \tau$ there exists for a lift $\tilde{p} \in \tilde{\Lambda}$ $(\pi(\tilde{p}) = p)$ a unique element $g \in G$ with $g\tilde{p} = \tilde{\varphi}_{\lambda(\tau)}\tilde{p} \in \tilde{\Lambda}$. We call the conjugacy class $[g]$ in G the Frobenius class of τ, which we denote by $[\tau]$. We have the following analogue of Theorems 8.5 and 8.6.

THEOREM 9.5. (Chebotarov theorem for hyperbolic flows.) *Let* $\tilde{\varphi}_t : \tilde{\Lambda} \to \tilde{\Lambda}$ *be a Galois G-extension of a hyperbolic flow* $\varphi_t : \Lambda \to \Lambda$.

(a) *If* $\tilde{\varphi}$ *is weak-mixing then for every continuous class function* $F \in C(G,\mathbb{C})$ (i.e. $F(hgh^{-1}) = F(g)$, *for all* $h \in G$) *we have*

$$\frac{1}{\pi'(t)} \sum_{\lambda(\tau)\le t} F([\tau]) \to \int F(g)dg, \text{ as } t \to \infty .$$

(b) *If* $\varphi,\tilde{\varphi}$ *are topologically transitive and* G *is finite then for every class function* $F \in C(G,\mathbb{C})$ *we have*

$$\frac{1}{\pi'(t)} \sum_{\lambda(\tau) \leq t} F([\tau]) \to \int F(g)dg, \text{ as } t \to \infty .$$

(In part (b) the condition that φ is weak-mixing forces both $\tilde{\varphi}$ and the associated G-extension of the suspended flow $\tilde{\sigma}_f$ to be weak-mixing. In part (a) $\tilde{\varphi}$ is weak-mixing if and only if $\tilde{\sigma}_f$ is weak-mixing, cf. [15] for details.)

Applications to geodesic flows: In the three distribution theorems given above we have dealt with hyperbolic flows. A special case is the geodesic flow $\varphi_t : M \to M$ on the unit tangent bundle of a compact manifold V with strictly negative sectional curvatures. These geodesic flows are always weak-mixing. There is then a one-one correspondence between the closed orbits and the (directed) closed geodesics. (The least periods of the orbits being the lengths of the closed orbits.)

Thus $\pi'(x)$ is the number of closed geodesics on V of least period at most x, and Theorem 9.3 gives $\pi'(x) \sim e^{hx}/hx$. (This result for closed geodesics was originally proved by Margulis [57].)

Let $F : V \to \mathbb{R}$ be a continuous function. If $\rho : M \to V$ is the natural projection of the unit tangent bundle onto the manifold, then we can apply Theorem 9.4 to $F \circ \rho$ to deduce:

$$\frac{1}{\pi'(t)} \sum_{\ell(\gamma)\leq x} \int_\gamma F d\ell \to \int F d(\rho^* m)$$

where γ denotes a (directed) closed geodesic of least period $\ell(\gamma), \int_\gamma F d\ell$ denotes the integral of F around the geodesic γ with respect to one-dimensional Lebesgue measure, and $\rho^* m$ is the projection of the measure of maximal entropy.

Finally, let M^k denote the bundle of k-dimensional (oriented) frames over V. (The special case $M^1 = M$ is the unit tangent bundle.) We define an extension $\tilde{\varphi}_t : M^k \to M^k$ of $\varphi_t : M \to M$ by parallel transport of the frame around the geodesic associated to the orbit of φ. The flow $\tilde{\varphi}$ is a $G = SO(n-1)$ extension of $\varphi_t : M \to M$, where n is the dimension of V. Each closed geodesic γ gives rise to a conjugacy class $[\gamma]$ in G corresponding to the holonomy group (i.e. the elements of G which come from parallel translations of a frame around γ). From Theorem 9.5 we see that whenever $\tilde{\varphi}$ is weak-mixing we have, for any continuous class function $F \in C(G,\mathbb{C})$:

$$\frac{1}{\pi'(t)} \sum_{\ell(\gamma)\leq x} F([\gamma]) \to \int F dg\,.$$

There are geometric criteria on V for the associated frame flow $\tilde{\varphi}$ to be weak-mixing (cf. [20] for example).

Application to twisted orbits. Let $\varphi_t : \Lambda \to \Lambda$ be a hyperbolic flow on a basic set.

We denote the k-dimensional frames corresponding to E^u_x (where $k = \dim E^u$) by F and write $F = A \oplus B$, according to the orientation of the frames. We say that a closed orbit τ is *twisted* if $D\varphi_{\lambda(\tau)} A_x = B_x$ (and $D\varphi_{\lambda(\tau)} B_x = A_x$) for $x \in \tau$.

There is a natural $\mathbb{Z}_2$-extension $\tilde{\varphi}_t : \tilde{\Lambda} \to \tilde{\Lambda}$ of φ_t defined by the effect of $D\varphi_t$ on the orientation of E^u.

Clearly, $\tilde{\varphi}$ is topologically weak-mixing if E^u is *not* orientable. Theorem 9.5 (b) then shows that closed orbits are equally distributed between those which are twisted and those which are not.

Notes

Lemmas 9.1 and 9.2 appear in the work of Bowen [15].

Zeta functions for Axiom A flows were introduced in [95] based on the Selberg zeta function for geodesic flows. The definition we give follows Ruelle's [80] which is a close analogue of the Riemann zeta-function.

Theorem 9.1 (a) appears in [15] and Theorem 9.1 (b) was proved in [66] as a generalisation of a result in [80].

Theorem 9.2 can be deduced from [67] and [71].

Theorem 9.3 was proved by Parry and Pollicott. Earlier asymptotic results for geodesic flows were proved in [42] and [57]. Less precise growth rates were proved in [92] and [15].

Theorem 9.4 was first proved in [14]; the proof we present may be found in [67].

Theorem 9.5 is due to Parry and Pollicott [68]. This work was motivated by the earlier work of Adachi and Sunada [3].

The application to frame flows is described in [68].

CHAPTER 10

FURTHER EXTENSION RESULTS

We have now completed the derivation of the main theorems concerning the asymptotic formulae for closed orbits. For their proofs it was sufficient for us to know the analytic domain of the appropriate zeta-functions in a neighbourhood of the line $\mathcal{R}(s) = h$. However, with only a little more work one can prove a much stronger result on the meromorphic domain of zeta-functions (although in general this will not lead to additional asymptotic results). The original extension (Theorem 5.6) required relatively weak estimates on the spectrum of the Ruelle operator (Theorem 4.5). Our improved estimates will result from a more detailed analysis of this spectrum. Using the notation of Chapter 4 we shall consider the Ruelle operator $L_f : F_\theta^+ \to F_\theta^+$ corresponding to $f = u+iv \in F_\theta^+$.

Let $T : B \to B$ be a bounded linear operator on a Banach space B. The *essential spectrum* esp (T) of T consists of those $\lambda \in sp(T) \subseteq \mathbb{C}$ for which any of the following three equivalent statements is valid:

(i) Range $(\lambda I-T)$ is not closed in B

(ii) λ is a limit point of $sp(T)$

(iii) $\bigcup_{r=1}^{\infty} \ker(\lambda I-T)^r$ is infinite dimensional (cf. Browder [21]). In particular, $\mathrm{esp}(T)$ does not contain isolated eigenvalues of finite multiplicity.

The *essential spectral radius* is defined to be $\rho_e(T) = \sup \{|\lambda| : \lambda \in \mathrm{esp}(T)\}$. There is a useful formula for $\rho_e(T)$ due to Nussbaum [60]:

PROPOSITION 10.1. $\rho_e(T) = \lim_{n\to\infty} \|T^n\|_e^{1/n}$, *where* $\|T\|_e = \inf \{\|T-K\| : K : B \to B$ *is a compact operator*$\}$.

Our first result deals with the spectrum of L_f.

THEOREM 10.2. *For* $f = u+iv \in F_\theta^+$ *the spectrum of* $L_f : F_\theta^+ \to F_\theta^+$ *consists of two distinct parts:*

(i) *The essential spectrum consisting of the whole disc* $\{z \in \mathbb{C} : |z| \le \theta e^{P(u)}\}$.

(ii) *Isolated eigenvalues (of finite multiplicity) contained in the annulus* $\{z \in \mathbb{C} : \theta e^{P(u)} < |z| \le e^{P(u)}\}$.

PROOF. We can assume L_u is normalised (i.e. $e^{P(u)} = 1$). We begin by showing that $\rho_e(L_f) \leq \theta$. For $n \geq 1$ let C denote a cylinder of the form $C = \{z \in X_A^+ \mid z_i = x_i, 0 \leq i \leq n-1\}$, where $A(x_i, x_{i+1}) = 1, 0 \leq i \leq n-2$. We can write $|C| = n$, and associate to each cylinder C an element $x_C \in C \subseteq X_A^+$. (The specific choice made will be unimportant.)

We define an operator $E_n : F_\theta^+ \to F_\theta^+, n \geq 1$, by $E_n(f)(x) = \sum_{|C|=n} \chi_C(x).f(x_C)$, where the sum here is clearly finite.

We can make the following explicit estimates:

(i) $|f-E_n(f)|_\infty \leq \sup_{|C|=n} \left\{ \sup_{x \in C} |f(x)-f(x_C)| \right\} \leq |f|_\theta \, \theta^n$

(ii) $|f-E_n(f)|_\theta \leq |f|_\theta$

(since for $k \geq n$, $var_k(f-E_n(f)) = var_k(f)$ and for $k \leq n$, $var_k(f-E_n(f)) \leq |f|_\theta \, \theta^n$, by estimate (i)).

The operator $E_n : F_\theta^+ \to F_\theta^+$ has a finite dimensional range, and hence is compact. Therefore $K_n = L_f^n E_n : F_\theta^+ \to F_\theta^+$, $n \geq 1$, is also compact, since the composition of a compact operator with a bounded linear operator is again compact. By the basic inequality (Proposition 2.1) we have

$$\| (L_f^n - K_n) h \|_\theta = \| L_f^n (h - E_n(h)) \|_\theta \leq (C+1) | h - E_n(h) |_\infty + \theta^n | h - E_n(h) |_\theta$$
$$\leq (C+1) |h|_\theta \theta^n + |h|_\theta \theta^n \leq (C+2) \|h\|_\theta \theta^n .$$

Thus by Proposition 10.1 we have $\rho_e(L_f) \leq \lim_{n \to \infty} \| L_f^n - K_n \|^{1/n} \leq \theta$.

Part (ii) of the theorem follows from the definition of essential spectrum and our estimate $\rho_e(L_f) \leq \theta$. Assume that $|\lambda| < \theta$ and choose $h \in F_\theta^+$ such that $h \neq 0$ but $L_u h = 0$. We can define $g \in F_\theta^+$ by $g = \sum_{n=0}^{\infty} e^{-iv^n} h \circ \sigma^n \lambda^n$ and observe that $L_f(g) = \lambda g$.

Thus for a judicious choice of h with $g \not\equiv 0$ we have λ lies in the (essential) spectrum of L_f. By the compactness of the spectrum we see that $|z| = \theta$ also lies in the (essential) spectrum. This completes the proof.

This result on the spectrum leads to a meromorphic extension of the zeta-function $\zeta(f) = \exp \sum_{n=1}^{\infty} \frac{1}{n} \sum_{\sigma^n x = x} e^{f^n(x)}$. The principle is essentially the same as that in Chapter 5. However, since we have very precise information on the spectrum of the Ruelle operator from Theorem 10.2 it is appropriate to translate this into an extension result for $\zeta(f)$ with as much care as possible.

To prepare for the main result of this chapter we choose $\rho > \theta$ such that L_f

has no eigenvalues of modulus ρe^{P} where $P = P(\mathcal{R}f)$. Clearly ρ can be chosen as close to θ as we wish, and there are only finitely many eigenvalues (each with finite multiplicity) of modulus greater than ρe^{P}. Moreover the projection Q onto the part of the spectrum of L_f inside $\{z : |z| \leq \rho e^{P}\}$ commutes with L_f and QL_f has spectral radius strictly less than ρe^{P}.

Evidently L_f can be expressed as

$$L_f w = \sum_{i=1}^{N} \lambda^{(i)} \sum_{\alpha,\beta} v_{\alpha}^{(i)} L_{\alpha,\beta}^{(i)} \mu_{\beta}^{(i)} w + QL_f w \tag{10.1}$$

where $\lambda^{(i)}$ $(i = 1,\ldots,N)$ are the eigenvalues of L_f with modulus greater than ρe^{P}, $v_{\alpha}^{(i)}$ belong to the corresponding eigenspaces, $\mu_{\beta}^{(i)}$ belong to the dual eigenspaces and $\mu_{\beta}^{(j)}\left(v_{\alpha}^{(i)}\right) = \delta_{i,j}\delta_{\alpha,\beta}$. The matrices $L^{(i)}$, of course, are in Jordan normal form and the multiplicity of the eigenvalue $\lambda^{(i)}$ is $\nu^{(i)} = \text{trace } L^{(i)}$. Iterating (10.1) we have

$$L_f^m w = \sum_{i=1}^{N} \lambda^{(i)m} \sum_{\alpha,\beta} v_{\alpha}^{(i)} L_{\alpha,\beta}^{(i)m} \mu_{\beta}^{(i)} w + QL_f^m w .$$

Fixing $m > 0$, we denote by $\sum_{\eta}^{*}$ the sum over all permissible words of length m such that the periodic concatenation $\eta^* = \eta \vee \eta \vee \ldots$ is permissible and we denote by $\sum_{\eta}$ the sum over all permissible words of length m. Define $\eta^{\#} = \eta^*$ when η^* is permissible and otherwise define $\eta^{\#}$ arbitrarily subject to $\eta \vee \eta^{\#}$ being permissible. Finally we let the word η stand for the cylinder it defines so that χ_{η} is the characteristic function of the set of points which begin with η.

We shall be especially interested in

$$\zeta_m = \sum_{\mathrm{Fix}_m} e^{f^m(x)} = \sum_{\eta}^{*} \exp f^m(\eta^*).$$

Notice that

$$\begin{aligned} L_f^m \chi_{\eta}(x) &= \exp f^m(\eta \vee x) \quad \text{if } \eta \vee x \text{ is permissible} \\ &= 0 \qquad \text{otherwise,} \end{aligned} \tag{10.2}$$

and therefore

$$\begin{aligned} \zeta_m &= \sum_{\eta} \left(L_f^m \chi_{\eta}\right)(\eta \vee \eta^{\#}) \\ &= \sum_{\eta} \Big[\sum_{i=1}^{N} \lambda^{(i)m} \sum_{\alpha,\beta} v_{\alpha}^{(i)}(\eta \vee \eta^{\#})\, L_{\alpha,\beta}^{(i)m} \mu_{\beta}^{(i)}(\chi_{\eta}) + \left(QL_f^m \chi_{\eta}\right)(\eta \vee \eta^{\#})\Big] \end{aligned}$$

$$= \sum_{i=1}^{N} \lambda^{(i)m} \sum_{\alpha,\beta} L_{\alpha,\beta}^{(i)m} \mu_\beta^{(i)} \left(\Sigma_\eta v_\alpha^{(i)} (\eta \vee \eta^\#) \chi_\eta \right) + \Sigma_\eta \left(Q L_f^m \chi_\eta \right) (\eta \vee \eta^\#)$$

$$= \zeta_m^{(0)} + \zeta_m^{(1)} + \zeta_m^{(2)} \text{ where}$$

$$\zeta_m^{(0)} = \sum_{i=1}^{N} \nu^{(i)} \lambda^{(i)m}$$

$$\zeta_m^{(1)} = \sum_{i=1}^{N} \lambda^{(i)m} \sum_{\alpha,\beta} L_{\alpha,\beta}^{(i)m} \mu_\beta^{(i)} \left(\Sigma_\eta v_\alpha^{(i)} (\eta \vee \eta^\#) \chi_\eta - v_\alpha^{(i)} \right)$$

$$\zeta_m^{(2)} = \Sigma_\eta \left(Q L_f^m \chi_\eta \right) (\eta \vee \eta^\#) .$$

Our aim is to estimate $\quad \zeta_m = \zeta_m^{(0)} + \zeta_m^{(1)} + \zeta_m^{(2)}$.

LEMMA 10.3. *There is a constant such that*

$$| \zeta_m^{(1)} | \leq \text{const.} \, (\rho e^P)^m .$$

PROOF. One first verifies

$$L_f^* \mu_\alpha^{(i)} = \lambda^{(i)m} \sum_\beta L_{\alpha,\beta}^{(i)m} \mu_\beta^{(i)}$$

from which we obtain

$$\zeta_m^{(1)} = -\sum_{i=1}^{N} \lambda^{(i)m} \sum_{\alpha,\beta} L_{\alpha,\beta}^{(i)m} \mu_\beta^{(i)} \left(v_\alpha^{(i)} - \Sigma_\eta v_\alpha^{(i)} (\eta \vee \eta^{\#}) \chi_\eta \right)$$

$$= -\sum_{i=1}^{N} \sum_{\alpha} \mu_\alpha^{(i)} \left(L_f^m \left(v_\alpha^{(i)} - \Sigma_\eta v_\alpha^{(i)} (\eta \vee \eta^{\#}) \chi_\eta \right) \right)$$

$$= -\sum_{i=1}^{N} \sum_{\alpha} \mu_\alpha^{(i)} \left(L_f^m v_\alpha^{(i)} - K_m v_\alpha^{(i)} \right)$$

where K_m is of finite rank and since $\| L^n - K_n \| \leq$ const. $(\rho e^P)^n$ the lemma is proved.

LEMMA 10.4. *There is a constant such that* $|\zeta_m^{(2)}| \leq$ const. $m(\rho e^P)^m$.

PROOF. Let $\bar{\eta}(m-k)$, $\eta(k)$ denote the words formed from the first $m-k$ symbols and the last k symbols of the word η of length m, respectively, so that $\eta = \bar{\eta}(m-k) \vee \eta(k)$. We need to assign, once and for all, a sequence to follow (permissibly) each symbol i - clearly $i^{\#}$ will do. Moreover for every η of length m and for every $1 \leq k \leq m$, $\eta(k) \vee \eta(1)^{\#}$ is permissible.

We define for $1 \leq k \leq m$ the functions

$$X_{\eta(k)}(x) = \exp - f^k(\eta(k) \vee \eta(1)^{\#}).\left(L_f^k \chi_{\eta(k)}\right)(x)$$

and for $2 \leq k \leq m$

$$\begin{aligned} Y_\eta(x) &= X_{\eta(k)}(x) - X_{\eta(k-1)}(x) \\ &= X_{\eta(k-1)}(x)\,\{\exp[f(\eta(k)\vee x) - f(\eta(k)\vee \eta(1)^\#)] - 1\} \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \eta(k)\vee x \text{ is permissible} \\ &= 0 \text{ otherwise}. \end{aligned}$$

We also define $Y_{\eta(1)} = X_{\eta(1)}$.

It is easy to show that

$$\| X_{\eta(k)} \|_\theta \le \text{const. and } \| Y_{\eta(k)} \|_\theta \le \text{const. } \theta^k$$

where the constants are independent of m, η, k.

Evidently

$$\sum_\eta \exp f^m(\eta \vee \eta(1)^\#).(QX_\eta)(\eta \vee \eta^\#) = \sum_\eta \left(QL_f^m \chi_\eta\right)(\eta \vee \eta^\#) = \zeta_m^{(2)},$$

and hence

$$\begin{aligned} \zeta_m^{(2)} &= \sum_\eta \exp f^m(\eta \vee \eta(1)^\#).(QX_\eta)(\eta \vee \eta^\#(1)) \\ &+ \sum_\eta \exp f^m(\eta \vee \eta(1)^\#)\,\{(QX_\eta)(\eta \vee \eta^\#) - (QX_\eta)(\eta \vee \eta(1)^\#)\} \\ &= A + B, \text{ say.} \end{aligned}$$

Notice that

$$A = \sum_{\eta} \exp f^m(\eta \vee \eta(1)^{\#}) \left\{ \sum_{k=2}^{m} (QX_{\eta(k)} - QX_{\eta(k-1)}) + QX_{\eta(1)} (\eta \vee \eta(1)^{\#}) \right\}$$

$$= \sum_{k=1}^{m} \sum_{\eta} \exp f^m(\eta \vee \eta(1)^{\#}).(QY_{\eta(k)}) (\eta \vee \eta(1)^{\#})$$

$$= \sum_{k=1}^{m} \sum_{\eta(k)} \sum_{\bar{\eta}(m-k)} \exp f^m (\bar{\eta}(m-k) \vee \eta(k) \vee \eta(1)^{\#}).(QY_{\eta(k)})(\bar{\eta}(m-k) \vee \eta(k) \vee \eta(1)^{\#}),$$

and also notice that

$$\sum_{\bar{\eta}(m-k)} \exp f^{m-k} (\bar{\eta}(m-k) \vee \eta(k) \vee \eta(1)^{\#}).(QY_{\eta(k)})(\bar{\eta}(m-k) \vee \eta(k) \vee \eta(1)^{\#})$$

$$= L_f^{m-k}(QY_{\eta(k)})(\eta(k) \vee \eta(1)^{\#}) .$$

So we conclude that

$$A = \sum_{k=1}^{m} \sum_{\eta(k)} \exp f^k(\eta(k) \vee \eta(1)^{\#}).\left(L_f^{m-k}(QY_{\eta(k)})(\eta(k) \vee \eta(1)^{\#}\right)$$

and

$$A \leq \sum_{k=1}^{m} \sum_{\eta(k)} \exp \mathcal{R}f^k(\eta(k) \vee \eta(1)^{\#}).\| L_f^{m-k} Q \| \|Y_{\eta(k)} \|_\theta \leq \text{const. } m(\rho e^P)^m .$$

On the other hand we have

$$|B| \le \sum_{\eta} \exp \mathcal{R}f^m(\eta \vee \eta(1)^*).|(QX_\eta)(\eta \vee \eta^*) - (QX_\eta)(\eta \vee \eta(1)^*)|$$

$$\le \sum_{\eta} \exp \mathcal{R}f^m(\eta \vee \eta(1)^*) \|Q\| \|X_\eta\|_\theta \theta^m \le \text{const.} (\theta e^P)^m.$$

The lemma is therefore proved.

We are now in a position to prove

THEOREM 10.5. *For* $f \in F_\theta^+$ *and*

$$Z(z,f) = \sum_{n=1}^{\infty} \frac{z^n}{n} \sum_{Fix_n} \exp f^n(x) = \sum_{n=1}^{\infty} \frac{z^n}{n} \zeta_n$$

the function $\exp - Z(z,f)$ *extends to an analytic function in the region of* $\mathbb{C} \times F_\theta^+$ *where* $\theta\,|z| < e^{-P(\mathcal{R}f)}$.

PROOF. Let ν_i be the multiplicity of the eigenvalue λ_i where $|\lambda_i| > \rho e^P > \theta e^P$, $i = 1,\dots,N$ then

$$\sum_{n=1}^{\infty} \frac{z^n}{n}\left(\zeta_n - \zeta_n^{(0)}\right) = \sum_{n=1}^{\infty} \frac{z^n}{n}\zeta_n - \sum_{n=1}^{\infty} \frac{z^n}{n} \sum_{i=1}^{N} \nu^{(i)}\lambda_i^n = \sum_{n=1}^{\infty} \frac{z^n}{n}\left(\zeta_n^{(1)} + \zeta_n^{(2)}\right)$$

and by lemmas 10.3 and 10.4 the latter series converges when $|z|\,\rho e^P < 1$. On the other hand, $\exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \zeta_n^{(0)}\right)$ extends to the analytic function $\prod_{i=1}^{N}(1-z\lambda_i)^{\nu_i}$. We conclude therefore that

$$\exp - Z(z,f).\prod_{i=1}^{N}(1-z\lambda_i)^{-\nu_i} \equiv \varphi(z,f)$$

extends to an analytic function in the region $|z|\,\rho e^{P(\mathcal{R}f)} < 1$. In other words, $\varphi(z,f)\prod_{i=1}^{N}(1-z\lambda_i)^{\nu_i}$ is the required extension when $|z|\,\rho e^{P(\mathcal{R}f)} < 1$. Finally we observe that we may choose $\rho > \theta$ as close to θ as we please. That $\varphi(z,f)\prod_{i=1}^{N}(1-z\lambda_i)^{\nu_i}$ is analytic follows from Appendix V.

At the end of Chapter 5 we gave an application of our original extension (theorem 5.6) to the zeta-function $\zeta_{-f}(s) = \exp \sum_{n=1}^{\infty} \frac{1}{n} \sum_{\mathrm{Fix}_n} e^{-sf^n(x)}$, where $s \in \mathbb{C}$, for $f \in F_\theta^+$ with $f > 0$ and $P(-f) = 0$. We can improve this result by using Theorem 10.5 as follows:

COROLLARY 10.6. $\zeta_{-f}(s)$ *has a non-zero meromorphic extension to the half-plane* $\mathcal{R}(s) > 1-\varepsilon$, *where* $\varepsilon > 0$ *is given by the identity* $P(-(1-\varepsilon)f) = |\log\theta|$.

(We recall that since $f > 0$, $t \mapsto P(-tf)$ is a strictly monotonic descreasing homeomorphism of the real line, and clearly $|\log \theta| > 0$.)

We can give a fairly simple example to show that the meromorphic extension in the above corollary (and hence the theorem) is essentially sharp with respect to the given data.

EXAMPLE 1. Given $\varepsilon > 0$ and $0 < \theta < 1$ we can construct $\sigma : X_A \to X_A$ and $f \in F_\theta^+$ with $f > 0$ such that $\zeta_{-hf}^{(s)}$ has an essential singularity at $s = s_0 \in \mathbb{R}$ with $|\log \theta| < P(-s_0 hf) \leq |\log \theta| + \varepsilon$.

Let $\sigma : X_n \to X_n$ be the full shift on n-symbols, i.e. $X_n = \prod_0^\infty \{1,2,\ldots,n\}$. Let $\beta_m, \beta > 0$ $(m \geq 0)$ be such that $|\beta - \beta_m| \leq C\theta^m$ for some constant $C > 0$. We define $f \in F_\theta^+$ by:

$$f(x) = \begin{cases} \beta_m & \text{if } x_i \neq n, 0 \leq i \leq m-1, \text{ and } x_m = n. \\ \beta & \text{if } x_i \neq n, \text{ for all } i \geq 0. \end{cases}$$

For sufficiently small $\alpha > 0$ we can choose $\beta_m, \beta > 0$ as

$$\beta_m = \begin{cases} \log n - \alpha \log[(1+\theta^m/m)/(1+\theta^{(m-1)}/m-1)] & \text{if } m \geq 2 \\ \log n - \alpha \log(1+\theta) & \text{if } m = 1 \\ \log n & \text{if } m = 0 \end{cases}$$

and $\beta = \log n$.

Since $|\log[(1+\theta^m/m)/(1+\theta^{m-1}/m)]| \leq \frac{\theta^m}{m} + \frac{\theta^{m-1}}{m-1}$ this sequence satisfies the required condition.

Assume $\sigma^k z = z$ and that the gaps between the occurrence of the symbol n are $k_1,\ldots,k_r$ where $k_1+\ldots+k_r = N$ and then

$$f^k(z) = (k-N)\,\beta_0 + \sum_{i=1}^{r} (\beta_1 + \beta_2 + \ldots + \beta_{k_i}).$$

For $s \in \mathbb{C}$ we can define $g \in C(X_2)$ by:

$$g(x) = \begin{cases} -s\beta_m + \log(n-1) & , \ \text{if } x_i = 1 \ (0 \leq i \leq m-1), x_m = 2 \\ -s\beta_0 & , \ \text{if } x_0 = 2 \\ -s\beta + \log(n-1) & , \ \text{if } x_i = 1 \ \text{for all } i \geq 0. \end{cases}$$

We clearly have the simple identity

$$\sum_{\sigma^k x = x \in X_n} e^{-sf^k(x)} = \sum_{\sigma^k x = x \in X_2} e^{g^k(x)} . \tag{10.9}$$

For any $N > 0$ we can define a square matrix by

$$P_N = \begin{bmatrix} e^{-s\beta_0} & (n-1)e^{-s\beta_1} & \cdot & \cdot & \cdot & (n-1)e^{-s\beta_{N-1}} & (n-1)e^{-s\beta} \\ e^{-s\beta_0} & 0 & & \mathbf{0} & & \vdots & \vdots \\ 0 & (n-1)e^{-s\beta_1} & \cdot & & & \vdots & \vdots \\ \vdots & \mathbf{0} & & \cdot & \cdot & (n-1)e^{-s\beta_{N-1}} & (n-1)e^{-s\beta} \end{bmatrix}$$

so that $\sum_{\sigma^k x = x} e^{g^k(x)} = \text{trace}\,(P_N)^k$, provided $N > k$.

For $\mathcal{R}(s)$ large we can find the explicit expression:

$$\begin{aligned} 1/\zeta_{-f}(s) &= \exp - \sum_{k=1}^{\infty} \frac{1}{k} \sum_{\sigma^k x = x} e^{g^k(x)} \qquad \text{(using (10.9))} \\ &= \lim_{N\to\infty} \exp - \sum_{k=1}^{\infty} \frac{1}{k} \,\text{trace}\,(P_N)^k \\ &= \lim_{N\to\infty} \det (I - P_N) \\ &= \lim_{N\to\infty} (1-(n-1)e^{-s\beta}) \left(1 - \sum_{k=0}^{N-1} (n-1)e^{-s(\beta_0+\dots+\beta_k)}\right) - (n-1)Ne^{-s(\beta+\beta_0+\dots+\beta_{N-1})} \\ &= (1-(n-1)e^{-s\beta}) \left(1 - \sum_{k=1}^{\infty} (n-1)^k \, e^{-s(\beta_0+\dots+\beta_k)}\right) \end{aligned}$$

$$= (1-(n-1)/n^s)\left(1-\sum_{k=1}^{\infty} (n-1)^k / n^{s(k+1)} (1+\theta^k/k)^{\alpha s} - {}^1/n^s\right).$$

The entropy h corresponds to the first positive pole, i.e.

$$1 = \sum_{k=1}^{\infty} (n-1)^k / n^{h(k+1)} (1+\theta^k/k)^{\alpha h} + n^{-h}.$$

To find an extension for $\zeta_{-f}(s)$ it suffices to find an extension for

$$F(s) = \sum_{k=1}^{\infty} (n-1)^k / n^{s(k+1)} (1+\theta^k/k)^{\alpha s}.$$

We can write $F(s) = \frac{1}{n^s} \sum_{k=1}^{\infty} [(n-1)/n^s]^k [(1+\theta^k/k)^{\alpha s} - 1] + \frac{(n-1)}{n^{2s}(1-(n-1)/n^s)}$. For $0 < s \leq h$ we can bound $A \leq \frac{[(1+\theta^k/k)^{\alpha s} - 1]}{s\theta^k/k} \leq B$, $A, B > 0$ and so:

$$\frac{A}{n^h} \leq \frac{G(s)}{|\log(1-\theta(n-1)/n^s)|} \leq B, \text{ where } F(s) = sG(s) + \frac{(n-1)}{n^{2s}(1-(n-1)/n^s)}.$$

Consider $s = s_1$ such that $\theta(n-1)/n^{s_1} = 1$ then as $s \searrow s_1$ we have $|G(s)| \to +\infty$, but $(s-s_1)\,G(s) \to 0$. We conclude that each of these functions, but particularly $\zeta_{-f}(s)$ must have an essential singularity at $s_1 = \frac{\log(n-1)}{\log n} - \frac{|\log\theta|}{\log n}$.

Since $\sigma : X_n \to X_n$ has topological entropy $\log n$ it is clear from Abramov's theorem that $h(\sigma_f)$ approaches 1 as n increases. We let $h = h(\sigma_f)$, then $P(-hf) = 0$ and $\zeta_{-hf}(s)$ has an essential singularity at $s_0 = s_1/h$.

Thus $P(-s_0hf) = P(-s_1f) \le \log n - s_1(\log n - \alpha)$

$$= \log n - \left[\frac{\log(n-1)}{\log n} - \frac{|\log\theta|}{\log n}\right](\log n - \alpha)$$

$$= |\log\theta| + \log\left(\frac{n}{n-1}\right) + \frac{\alpha\log[(n-1)\theta]}{\log n}.$$

We can assume that n is sufficiently large that $(n-1)\,\theta > 1$ and $\log(n/(n-1)) < \varepsilon/2$ and then $\alpha > 0$ can be chosen sufficiently small that $\alpha\log[(n-1)\,\theta]/\log n < \varepsilon/2$. Thus $P(-s_0f) \le |\log\theta| + \varepsilon$, as required.

The results in this chapter show that the zeta functions always have a non-zero meromorphic extension beyond the region where they naturally converge to non-zero analytic functions. A natural question is to what extent is this extension also analytic. In particular, for the case of the zeta function $\zeta_{-hf}(s)$ (σ_f weak-mixing, $f > 0$) is the extension to $1 - \varepsilon < \mathcal{R}(s) \le 1$: Is $\zeta_{-hf}(s)$ analytic (except for the simple pole at $s = 1$)? The next example gives a negative answer.

Example 2. Let $X_2 = \prod_0^\infty \{1,2\}$ and consider $f : X_2 \to \mathbb{R}$ defined by:

$$f(x) = \begin{cases} \alpha & \text{if } x_0 = 1 \\ \beta & \text{if } x_0 = 2 \end{cases}$$

where $\alpha, \beta > 0$ and α/β is irrational. By direct computation we can see that

$$\sum_{\sigma^k x = x} e^{-sf^k(x)} = \operatorname{trace}(P(s))^k\,, \text{ where } P(s) = \begin{pmatrix} e^{-\alpha s} & e^{-\beta s} \\ e^{-\alpha s} & e^{-\beta s} \end{pmatrix}.$$

$$\text{Thus } \zeta_{-f}(s) = \exp \sum_{k=1}^{\infty} \frac{1}{k} \sum_{\sigma^k x = x} e^{-sf^k(x)} = \exp \sum_{k=1}^{\infty} \frac{1}{k} \operatorname{trace}(P(s))^k$$

$$= \frac{1}{\det(I-P(s))} = 1/(1-e^{-\alpha s} - e^{-\beta s})\,.$$

The topological entropy $h(\sigma_f)$ for σ_f can be determined from the position of the first positive pole for $\zeta_{-f}(s)$, i.e. $e^{-\alpha h} + e^{-\beta h} = 1$. The poles for $\zeta_{-hf}(s)$ are determined by the simple condition: $e^{-\alpha hs} + e^{-\beta hs} = 1$. We observe that the assumption α/β is irrational implies that $\zeta_{-hf}(s)$ has no other poles on $\mathcal{R}(s) = 1$ other than $s = 1$. In particular, we can deduce that σ_f is weak mixing (cf. Theorem 6.3).

Consider the trigonometric polynomial $Q(s) = e^{-\alpha hs} + e^{-\beta hs} - 1$. For $\varepsilon > 0$ we see that for any $1-\varepsilon < \sigma < 1$ the function $t \mapsto Q(\sigma+it)$ $(t \in \mathbb{R})$ takes values arbitrarily close to zero. Since $Q(s)$ is a complex almost periodic function this implies that $Q(s)$ has zeros arbitrarily close to any line $\mathcal{R}(s) = \sigma$ (or equivalently, $\zeta_{-hf}(s)$ has poles arbitrarily close to this line). (Cf. [24].)

In particular, we can conclude the following: The zeta function $\zeta_{-hf}(s)$ has poles at $s = \sigma_n + it_n$, where $\sigma_n < 1$ and $\sigma_n \nearrow 1$.

The corollary 10.6 and the two examples above all describe the domain of zeta-functions $\zeta_{-hf}(s)$ for suspended flows σ_f. In keeping with our general philosophy these can be used to give results for hyperbolic flows. In particular, given a weak-mixing hyperbolic flow we can reduce the analysis of its zeta function $\zeta(s)$ to that of appropriate zeta-functions for suspended flows (Appendix III). Conversely, to construct (counter) examples of hyperbolic flows it is frequently easier to construct first suspended flows with the appropriate properties and then to use the "embedding theorem" to construct corresponding examples of hyperbolic flows (Appendix III).

In view of these two approaches we can simply translate the corollary 10.6 and examples 1 and 2 from the context of suspended flows to that of hyperbolic flows to deduce the following result.

THEOREM 10.7.

(i) *Let* $\varphi_t : M \to M$ *be a weak-mixing hyperbolic flow (with topological entropy* $h > 0$) *then there exists* $\varepsilon > 0$ *such that* $\zeta(s)$ *has a non-zero meromorphic extension to* $\mathcal{R}(s) > 1-\varepsilon$.

(ii) *There exist examples of hyperbolic flows for which* $\zeta(s)$ *does not extend meromorphically to the entire complex plane.*

(iii) *There exist examples of hyperbolic flows for which* $\zeta(s)$ *does not extend analytically to the strip* $1-\delta < \mathcal{R}(s) < 1$, *for any* $\delta > 0$.

In parts (i) and (ii) of Theorem 10.7 it is difficult to give a useful quantitative estimate for the size of the extension, or the position of essential singularities. This is simply because the characterisations in the suspended flow setting do not conveniently translate into appropriate terms for hyperbolic flows.

Notes

The definition of the essential spectrum and essential spectral radius is originally due to Browder [21].

Proposition 10.1 is due to Nussbaum [60].

Theorem 10.2 is due to Pollicott [72]. A similar result for interval maps by Keller appears in [51].

Theorem 10.5 and its related lemmas and Corollary 10.6 are due to Haydn [36]. Our account is based on a version of Haydn's proof provided by Ruelle. Prior to Haydn's result there were weaker versions of this theorem due to Ruelle [82] and Pollicott [72].

Example 1 is taken from Pollicott's article [72], which in turn is a development of earlier examples of Gallavotti [34] and Pollicott [71].

Example 2 was discovered independently by Ruelle [83] and Pollicott [71].

Finally, Theorem 10.7 can be found in [36].

CHAPTER 11

ATTRACTORS AND SYNCHRONISATION

In the preceding chapters we were mainly concerned with results for general hyperbolic flows. As we shall see it is sometimes appropriate to take into account the ambiant manifold M.

A basic set Λ for a hyperbolic flow $\varphi_t : M \to M$ is called an *attractor* if there exists some open neighbourhood $U \supset \Lambda$ such that $\Lambda = \bigcap_{t \geq 0} \varphi_t U$.

This amounts to a strengthening of the hypothesis $\Lambda = \bigcap_{t \in \mathbb{R}} \varphi_t U$ in the definition of a basic set. The geodesic flow for a manifold with negative sectional curvatures is an example of a hyperbolic flow with $\Lambda = M$ so that with the choice $U = M$ it is clearly an attractor.

The complementary notion is that of a *repellor* Λ where one requires an open neighbourhood $U \supset \Lambda$ such that $\Lambda = \bigcap_{t \leq 0} \varphi_t U$. There is a trivial correspondence between repellors and attractors, in that if $\psi_t : \Lambda \to \Lambda$ is the hyperbolic flow given by $\psi_t = \varphi_{-t}$ (i.e. by reversing the time direction) then an attractor for φ is a repellor for ψ and vice versa.

In chapter 9 we described equilibrium states for Hölder continuous functions on basic sets. Of particular importance was the measure of maximal entropy (associated to the function $F = 0$).

For attractors there is a second, more geometric, canonical equilibrium state.

For each point $x \in \Lambda$ and $t > 0$ we can consider the map $D\varphi_t : T_xM \to T_{\varphi_t x}M$ and its restriction $D\varphi_t : E^u_x \to E^u_{\varphi_t x}$. We define $\lambda^u(x) = \lim_{t \to 0} \frac{1}{t} \log \operatorname{Det} (D\varphi_t | E^u_x)$, which we call the *expansion coefficient* of φ at $x \in \Lambda$. Thus we have a well-defined map $\lambda^u : \Lambda \to \mathbb{R}$. When φ_t is C^2 it is known that the splitting $x \mapsto E^u_x \oplus E^s_x$ is Hölder continuous (cf. [39]) and so we see that $x \mapsto \lambda^u(x)$ is Hölder continuous.

The value $\lambda^u(x)$ has an intuitive interpretation as the infinitesimal expansion along the unstable bundle as the point x moves along its orbit. We can associate to $-\lambda^u$ a unique equilibrium state m supported on Λ. This measure is called the *Sinai-Ruelle-Bowen measure (S.R.B.-measure).*

We define the *basin of attraction* $B(\Lambda) \subset M$ of an attractor $\varphi_t : \Lambda \to \Lambda$ by $B(\Lambda) = \{x \in M : d(\varphi_t x, \Lambda) \to 0 \text{ as } t \to +\infty\}$ (or equivalently, $B(\Lambda) = \bigcup_{x \in \Lambda} W^s(x)$). The following result gives the most important characterisation of m:

THEOREM 11.1. *For almost all* $x \in B(\Lambda)$, *with respect to a Riemannian volume on* M, *we have* $\frac{1}{T}\int_0^T F(\varphi_t x)dt \to \int F dm$ *as* $T \to +\infty$, *whenever* $F \in C^0(M)$.

PROOF.

Let $U \subset B(\Lambda)$ be an open set with $\bigcap_{t=0}^{\infty} \varphi_t U = \Lambda$ and let $F \in C^0(\Lambda)$. For any $\delta > 0$ let $E(\delta) = \{x \in U : \overline{\lim} \left| \frac{1}{T}\int_0^T F(\varphi_t x)dt - \int F dm \right| \geq \delta\}$. We want to show that the volume of $E(\delta)$ is zero, then the result is immediate.

We choose $\varepsilon > 0$ sufficiently small that for $d(x,y) \leq \varepsilon$ and $0 \leq t \leq 1$ we have $|F(\varphi_t x) - F(\varphi_t y)| < \delta/4$. If we define $C_n(\alpha) = \{x \in U : \left| \frac{1}{n}\int_0^n F(\varphi_t x)dt - \int F dm \right| \geq \alpha\}$, $\alpha > 0$, then

$$E(\delta) \subset \bigcap_{N=0}^{\infty} \bigcup_{n=N}^{\infty} C_n\left(\frac{3}{4}\delta\right) \tag{11.1}$$

and
$$\bigcap_{N=0}^{\infty} \bigcup_{n=N}^{\infty} C_n\left(\frac{3}{4}\delta\right) \subset E\left(\frac{3}{4}\delta\right) \quad . \tag{11.2}$$

We fix $N \geq 1$ and construct a family of finite subsets $S_n \subset C_n(\delta/2) \cap \Lambda$, $n \geq N$, such that S_n is a set of maximal cardinality satisfying:

(i) $B_x(\varepsilon,n) \cap B_y(\varepsilon,k) = \emptyset$, $x \in S_n$, $y \in S_k$, $N \le k < n$

(ii) $B_x(\varepsilon,n) \cap B_{x'}(\varepsilon,n) = \emptyset$, $x,x' \in S_n$, $x \neq x'$

where $B_x(\varepsilon,T) = \{y \in M : d(\varphi_t x, \varphi_t y) \le \varepsilon$, for all $0 \le t \le T\}$. Choose $\varepsilon > \alpha > 0$ such that $B_\Lambda(\alpha) = \{y \in M : d(y,\Lambda) < \alpha\} \subset U$.

LEMMA 11.2. $B_\Lambda(\alpha) \cap \bigcup_{n=N}^{\infty} C_n(\frac{3}{4}\delta) \le \bigcup_{k=N}^{\infty} \bigcup_{x \in S_k} B_x(2\varepsilon,k)$.

PROOF OF LEMMA 11.2. If $y \in B_\Lambda(\alpha) \cap C_n(\frac{3}{4} 4\delta)$, $n \ge N$, and $y \in W^s_\tau(\varepsilon)$, with $z \in \Lambda$, then $z \in C_n(\delta/4) \cap \Lambda$. By the maximality of $\{S_m\}$ we conclude that $B_z(\varepsilon,n) \cap B_x(\varepsilon,k) \neq \emptyset$ for some $x \in S_k$, $N \le k \le n$. Therefore,

$$y \in W^s_\tau(\varepsilon) \subset B_z(\varepsilon,n) \subset B_x(2\varepsilon,k).$$

The proof of the following technical lemma can be found in [17]:

VOLUME LEMMA. For every $\varepsilon > 0$ there exists $A = A(\varepsilon) > 1$ with

$$\frac{1}{A} \le \frac{\text{Volume}(B_x(\varepsilon,T))}{\exp - \int_0^T \lambda^u(\varphi_t x)dt} \le A, \qquad \text{for all } x \in \Lambda, T \ge 0.$$

Combining these two lemmas we see that:

$$\text{Volume } \left(B_\Lambda(\alpha) \cap \bigcup_{n=N}^{\infty} C_n(\tfrac{3}{4}\delta)\right)$$

$$\leq A \sum_{k=N}^{\infty} \sum_{x \in S_k} \exp - \int_0^k \lambda^u(\varphi_t x)dt. \tag{11.3}$$

We next define $V_N = \bigcup_{k=N}^{\infty} \bigcup_{x \in S_k} B_x(\varepsilon,k)$, where the union, as we have observed, is disjoint.

LEMMA 11.3. $\lim_{N \to \infty}$ *volume* $(V_N) = 0$.

PROOF. For $x \in S_k \subset C_k(\delta/2)$ we have $B_x(\varepsilon,k) \subset C_k(\delta/4)$, from the definition of $\varepsilon > 0$. Thus $V_N \subset \bigcup_{k=N}^{\infty} C_k(\delta/4)$. (11.4)

Since m is ergodic, $m(E(\delta/4)) = 0$ by the ergodic theorem. Thus

$$0 = m(E(\delta/4)) \geq m\left(\bigcap_{N=0}^{\infty} \bigcup_{k=N}^{+\infty} C_k(\delta/4)\right) \qquad \text{(by (11.2))}$$

$$= \lim_{N \to +\infty} m\left(\bigcup_{k=N}^{\infty} C_k(\delta/4)\right).$$

and the result follows from (11.4).

By the volume lemma we can write:

$$m(V_N) \geq A \sum_{k=N}^{\infty} \sum_{x \in S_k} \exp - \int_0^k \lambda^u(\varphi_t x)dt \to 0 \text{ as } N \to \infty,$$

and substituting directly into (11.3) we conclude that

$$\lim_{N \to \infty} \text{Volume } \left(B_\Lambda(\alpha) \cap \bigcup_{n=N}^{\infty} C_n\left(\tfrac{3}{4}\,\delta\right)\right) = 0\,.$$

Therefore, by (11.1) we have Volume $(B_\Lambda(\alpha) \cap E(\delta)) = 0$. To remove the $B_\Lambda(\alpha)$ we observe that:

(i) $\varphi_t E(\delta)$, for all $t \geq 0$

and (ii) $\varphi_t U \subseteq B_\Lambda(\alpha)$ for all sufficiently large t,

so we deduce that Volume $(E(\delta)) = 0$.

To complete the proof of the theorem, we need a uniform result for $F \in C^0(M)$. By a standard argument it suffices to take a countable dense family $\{F_n\} \leq C^0(M)$ and the union of the sets $E(\delta)$ still has zero volume.

If $\Lambda = M$ and there exists a φ-invariant measure ν which is equivalent to the volume then by comparing the above Theorem with the usual Birkhoff ergodic theorem we see that ν is the SRB-measure. Therefore, for the geodesic flow example the SRB-measure is precisely the Liouville measure.

We now have two canonical measures for C^2 attractors; the measure of maximal entropy μ, and the SRB-measure m. In certain special cases these two measures will coincide, but generically they will be different. When these two measures are the same we shall say that the flow is *synchronised.* A trivial sufficient condition for a flow on an attractor to be synchronised is that the function λ^u is constant.

EXAMPLE 1. Consider the hyperbolic toral automorphism

$$(x,y) + \mathbb{Z}^2 \to (2x+y, x+y) + \mathbb{Z}^2,$$

and let $r : \mathbb{R}^2/\mathbb{Z}^2 \to \mathbb{R}^+$ be a constant function. For the associated suspended flow $\varphi_t : M \to M$ the manifold M is an attractor for which λ^u is constant, and consequently the measure of maximal entropy and the SRB-measure M are both equal to the (normalised) Lebesgue measure on M.

EXAMPLE 2. Let S be a compact surface of constant negative curvature $K = -1$. Such surfaces have an associated geodesic flow $\varphi_t : M \to M$ on the unit tangent bundle M to S. This is a hyperbolic flow with $\Lambda = M$ (cf. Appendix IV). The algebraic presentation of this flow takes the form $M = G/\Gamma$ where $\pm I \in \Gamma \subset G = SL(2,\mathbb{R})$ and Γ is a co-compact subgroup. The flow is written $\varphi_t(g\Gamma) = g_t g\Gamma$, where $g_t = \mathrm{diag}\,(e^t, e^{-t})$ (cf. Appendix IV).

The unstable manifolds for φ are the orbits of the horocycle flow corresponding to the matrix $h_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$, $t \in \mathbb{R}$. By simple algebra we see $g_t h_s = h_{se^t} g_t$. Letting $s \to 0$ we have $\mathrm{Det}\,(D\varphi_t \mid E_u) = e^t$. Thus $\lambda^u(x) = \lim_{t \to 0} \frac{1}{t} \log (e^t) \equiv 1$. We conclude that $\lambda^u \equiv 1$ and the two canonical measures coincide, and are equal to the Liouville (or Haar) measure on M.

When the two canonical measures are *not* coincident it is possible to reparameterise the flow so that the resulting flow is synchronised.

Let $\alpha : \Lambda \to \mathbb{R}^+$ be a strictly positive continuous function and define a map $k : \Lambda \times \mathbb{R} \to \mathbb{R}$ by $k(x,t) = \int_0^t \alpha(\varphi_u x)du$. This map has a unique 'inverse' defined by $k(x,\ell(x,t)) = \ell(x,k(x,t)) = t$ for all $x \in M, t \in \mathbb{R}$. We define a new flow $\psi_t : \Lambda \to \Lambda$ by $\psi_t(x) = \varphi_{\ell(x,t)}(x)$. This flow has the same orbits in Λ as the old flow $\varphi_t : \Lambda \to \Lambda$, but with a different parameterisation. At each point $x \in \Lambda$ we have $D_x\psi = \alpha(x)^{-1} D_x\varphi$, by the chain rule. In particular, $\alpha(x)^{-1}$ represents the 'change in velocity' between the two flows.

For any φ-invariant probability measure ν there is a corresponding ψ-invariant probability measure ν' with $\nu \sim \nu'$ and such that $\frac{d\nu'}{d\nu}(x) = \frac{\alpha(x)}{\int \alpha d\nu}$. Using Abramov's theorem we can related the entropies of the measures ν and ν' (with respect to the flows φ and ψ, respectively) by $h(\psi,\nu') = h(\varphi,\nu)/\int \alpha d\nu$.

If τ is a closed orbit for φ of least period $\lambda(\tau)$ then it is again a closed orbit for ψ of ψ-least period $\int_0^{\lambda(\tau)} \alpha(\varphi_t x)dt$ $(x \in \tau)$.

PROPOSITION 11.4. *If a* C^k *hyperbolic flow* $\varphi_t : \Lambda \to \Lambda$ *is reparameterised by a* C^k *function* $\alpha : \Lambda \to \mathbb{R}^+$ *then the resulting* C^k *flow is also hyperbolic.*

PROOF. We define a function $z : \Lambda \times E^s \to \mathbb{R}$ by $z(x,\xi) = \alpha(x) \int_0^\infty \xi\left(\frac{1}{\alpha \circ \varphi_u}\right) du$, which converges to a continuous function since $\xi \in E_x^s$. We introduce a new bundle $F_x^s = \{\xi + z(x,\xi)\, E_x^0 : \xi \in E_x^s\}$. To show this bundle is invariant under ψ_t we want to show $D\psi_s F_x^s = F_{\psi_s x}^s$ or equivalently, $D\psi_s(\xi+z(x,\xi)\, E_x^0) = D\varphi_t\xi + z(\varphi_t x, D\varphi_t\xi)\, E_{\varphi_t x}^0$ where $t = \ell(x,s)$ and $s = k(x,t)$, as before.

To establish this identity it suffices to differentiate both sides, and the equality corresponds to $(\xi+zE_x^0)(\alpha) = \alpha z'$, where we differentiate with respect to $(x,\xi)\mapsto(\varphi_t x, D\varphi_t\xi)$. The validity of this equality comes from substituting the above definition of z. The $D\psi$ invariant bundle F^s can easily be seen to be uniformly contracting.

Similarly, it is possible to construct a $D\psi$-invariant bundle F^u which is uniformly expanding (for example, by repeating the above construction with φ_{-t} replacing φ_t). Therefore, $\psi_t : \Lambda \to \Lambda$ has the hyperbolic splitting $T_\Lambda M = E^0 \oplus F^s \oplus F^u$. The other properties required for ψ to be a hyperbolic flow follow immediately from those for φ.

We can relate the expansion coefficients $\lambda_\varphi^u, \lambda_\psi^u : \Lambda \to \mathbb{R}^+$ for the flows φ and ψ, respectively, using the above proof of the proposition as follows:

The maps $v \mapsto v + z(x,v)\, E_x^0 \xmapsto{D\psi_s} D\varphi_t v + z(\varphi x, D\varphi_t\xi)\, E_{\varphi_t x}^0 \mapsto D\varphi_r v$ have jacobians $G(x)$, $\mathrm{Det}\left(D\psi_s \,|\, F_x^u\right)$, $G(\varphi_t x)^{-1}$ respectively and the composition $v \mapsto D\varphi_t v$ has jacobian $\mathrm{Det}\left(D\varphi_t \,|\, E_x^u\right)$.

If we write $\varphi^u(x,t) = \int_0^t \lambda_\varphi^u(\varphi_u x)du = \log|\mathrm{Jac}\left(D\varphi_t \mid E_x^u\right)|$ and

$\psi^u(x,s) = \int_0^s \lambda_\psi(\psi_u x)du = \log|\mathrm{Jac}\left(D\psi_s \mid F_x^u\right)|$ then we deduce that

$$\varphi^u(x,t) = \psi^u(x,s) + g(x) - g(\varphi_t x),$$

with $g = \log|G(x)|$, and

$$\psi^u(x,s) = \varphi^u(x,\ell(x,s)) + g(\psi_s x) - g(x).$$

We return to the problem of reparameterising a flow so that the new flow is synchronised. It is useful to make the following standing assumptions: *The flow* $\varphi_t : M \to M$ *is of class* C^2 *and the hyperbolic splitting* $T_\Lambda M = E^0 \oplus E^s \oplus E^u$ *is* C^1 (and thus so is the map $\lambda_\varphi^u : \Lambda \to \mathbb{R}^+$).

We come to our main result on synchronised flows:

THEOREM 11.5. *For a hyperbolic flow* $\varphi_t : \Lambda \to \Lambda$ *on an attractor, with the above properties, reparameterising by* $\alpha(x) = K/\lambda^u(x)$ *for any* $K > 0$, *gives a* C^1 *synchronised flow.*

PROOF. With the choice $\alpha(x) = K/\lambda^u(x)$ we have $\psi^u(x,s) = \varphi^u(x,\ell(x,s)+g(\psi_s x)-g(x)$ $= Kt + g(\psi_t x) - g(x)$. In particular, the weighting associated to each closed orbit τ takes the form $\psi^u(x,\lambda(\tau)) = K\lambda(\tau)$ $(x \in \tau)$. However, if we lift this to a modelling suspension we see that the functions whose equilibrium states correspond to the measure of maximal entropy and the SRB measure have the same values around closed orbits. It follows by Livsic's theorem (Proposition 3.7) that these functions differ by a coboundary. Therefore, we conclude that the equilibrium states on the suspension and thus the canonical measures on Λ coincide. Thus the flow ψ is synchronised.

In chapter 9 we considered various distribution results for closed orbits τ relative to their least periods $\lambda(\tau)$. (In particular, the spatial distribution of these orbits is relative to the measure of maximal entropy.) However, we can now consider a parallel analysis where we replace the least period of τ by the weight $\varphi^u(\tau) = \int_0^{\lambda(\tau)} \lambda_\varphi^u(\varphi_t x)dt$, which represents the total expansion in the unstable direction around the orbit τ. By the preceding theorem we can also interpret it as the least period of the closed orbit τ relative to the new flow ψ, when $K = 1$. This latter interpretation allows us to reformulate some of the asymptotic formulae for hyperbolic flows (from chapter 9).

THEOREM 11.6. *Let* $\varphi_t : \Lambda \to \Lambda$ *be a hyperbolic flow on an attractor and let* $\pi^u(t) = \mathrm{Card}\,\{\tau : \varphi^u(\tau) \leq t\}$, *for* $t > 0$. *Then*

either $\qquad \pi^u(t) \sim e^t/t$

or $\qquad \pi^u(t) \sim \dfrac{e}{e-1}\,\dfrac{e^{[t]}}{[t]}$

(depending on whether ψ *is topologically weak-mixing).*

Here we have used the fact that $h(\psi) = 1$.

THEOREM 11.7. *Let* $\varphi_t : \Lambda \to \Lambda$ *be a hyperbolic flow on an attractor with SRB measure* μ .

$$\frac{\displaystyle\sum_{x < \lambda(\tau) \leq x + \varepsilon} \frac{\lambda_F(\tau)}{\lambda(\tau)} e^{-\varphi^u(\tau)}}{\displaystyle\sum_{x < \lambda(\tau) \leq x + \varepsilon} e^{-\varphi^u(\tau)}} \to \int F d\mu\,, \text{ as } x \to +\infty, \quad \text{for all } F \in C^0(\Lambda)\,.$$

This follows immediately from remarks at the end of Chapter 7.

The hypothesis that the splitting $E^0 \oplus E^s \oplus E^u$ is C^1 was largely made for convenience. It is satisfied for certain well-known examples, the most important being geodesic flows for negatively curved surfaces. However, as was observed by Plante, it is frequently false. In these cases many of the results of this chapter remain valid, except that we must deal with the possibility that $\psi : \Lambda \to \Lambda$ is only Hölder continuous, which proves to be only a technical complication.

Notes

The SRB measure originated in the work of Sinai on Anosov systems and Gibbs measures in 1972 [94]. Ruelle extended Sinai's work to the context of Axiom A attracting diffeomorphisms [81]. Finally, Bowen and Ruelle developed the parallel theory for Axiom A attracting flows in 1975 [17]. It is in this last article that the proof of Proposition 11.1, including a proof of the omitted 'Volume Lemma' can be found.

The idea of synchronisation may be found in [69].

A good reference for the algebraic definition of the geodesic and horocycle flows is Cornfeld, Fomin and Sinai [25].

The proof of Proposition 11.4 for Anosov flows is due to Anosov and Sinai [6]. The modifications for the general case appear in the appendix to [69].

The distribution results for weighted orbits are due to Parry [69].

CHAPTER 12

CHEBOTAREV THEOREMS
FOR SOME NON-COMPACT GALOIS EXTENSIONS

In Chapter 8 we considered equidistribution results for compact Galois extensions of flows of finite type and in Chapter 9 we applied these results to compact Galois extensions of hyperbolic flows. Here we shall consider certain *non-compact* Galois extensions.

The canonical case will be a Galois extension by $\mathbb{Z}^d$, $d \geq 1$. Let $\sigma : X_A \to X_A$ be an aperiodic subshift of finite type. Consider the extension $\tilde{\sigma} : X_A \times \mathbb{Z}^d \to X_A \times \mathbb{Z}^d$ defined by $\tilde{\sigma}(x,z) = (\sigma x, g(x) + z)$, $z \in \mathbb{Z}^d$, where $g : X_A \to \mathbb{Z}^d$ is a locally constant function. Without loss of generality we may assume g is a function of two variables, i.e. $g(x) = g(x_0, x_1)$.

Since $\mathbb{Z}^d$ is an abelian group we need only consider the case of one-dimensional representations (i.e. characters) $\chi : \mathbb{Z}^d \to \{z \in \mathbb{C} : |z| = 1\}$. Following the approach in chapter 8 we introduce an L-function associated to the suspended flow $\sigma_{f,t} : X_f \to X_f$ defined by a strictly positive function $f \in F_\theta$.

We define

$$L(s,\chi) = \exp \sum_{n=1}^{\infty} \frac{1}{n} \sum_{\sigma^n x = x} \chi(g^n(x))\, e^{-shf^n(x)} \tag{12.1}$$

where $f^n(x) = f(x) + f(\sigma x) + \cdots + f(\sigma^{n-1}x)$, $g^n(x) = g(x) + g(\sigma x) + \cdots + g(\sigma^{n-1}x)$ and $h = h(\sigma_f)$ is the topological entropy of the flow. The complex function (12.1) can be seen to converge to a non-zero analytic function for $\mathcal{R}(s) > 1$ by comparison with the usual zeta-function (cf. Chapter 6).

Since $\chi \circ g : X_A \to \mathbb{C}$ satisfies $|\chi \circ g| = 1$ we can write $\chi \circ g(x) = e^{2\pi i k(x)}$, for some function $k \in F_\theta$ (determined up to an element of $C(X_A,\mathbb{Z})$. By Proposition 1.2 we can replace f by a function in F_θ^+ which differs, at most, by a coboundary, and in particular defines the same L-function. (We shall maintain the same notation.)

We write $L(s,\chi) = \zeta(-shf + 2\pi ik)$, where ζ is the general zeta-function studied in Chapter 5, and by theorem 5.6 we note the following:

PROPOSITION 12.1. $L(s,\chi)$ *has a non-zero analytic extension to a neighbourhood of* $\mathcal{R}(s) \geq 1$ *except for poles* $s = 1+it$ *where* $k - \frac{thf}{2\pi}$ *is cohomologous to an element of* $C(X_A,\mathbb{Z})$.

The characters $\chi : \mathbb{Z}^d \to \mathbb{C}$ are elements of the torus dual space $\hat{\mathbb{Z}}^d = \mathbb{R}^d/\mathbb{Z}^d$ of $\mathbb{Z}^d$. The map $\chi \mapsto L(s,\chi)$ can easily be seen to be analytic (on the torus) using the identity $L(s,\chi) = \zeta(-shf+2\pi ik))$ and Theorem 5.6. When we take $\chi = \chi_0$ to be the principal (trivial) character then $L(s,\chi_0) = \zeta(s)$ reduces to the usual zeta-function. Since $\zeta(s)$ has a simple pole at $s = 1$ we see that $(s,\chi) \mapsto L(s,\chi)$ is singular at $(1,\chi_0)$.

To proceed we shall make two assumptions about the domain of $L(s,\chi)$ which we shall later justify in certain cases.

(I) $L(s,\chi)$ is analytic at $(1+it,\chi) \neq (1,\chi_0)$, $\mathcal{R}(s) \geq 1$.

(II) In a neighbourhood of $(1,\chi_0)$ the pole $s = s(\chi)$ for $L(s,\chi)$ (with $s(\chi_0) = 1$) is smooth as a function of χ and $\nabla s(\chi)\,|_{\chi=\chi_0} = 0$ and $\det.\nabla^2\ \mathcal{R}(s(\chi))\,|_{\chi=\chi_0} < 0$, where ∇ and ∇^2 denote the gradient and the Hessian matrix, respectively.

The logarithmic derivative of $L(s,\chi)$ is

$$L'(s,\chi)/L(s,\chi) = -\sum_{k=1}^{\infty}\frac{1}{k}\sum_{\sigma^k x = x}\chi(g^k(x))\,hf^k(x)\,e^{-shf^k(x)}. \tag{12.2}$$

If $\alpha,\beta \in \mathbb{Z}^d$ then we have an orthonormality relation:

$$\int_{\mathbb{R}^n/\mathbb{Z}^n}\chi(-\alpha)\chi(\beta)\,d\chi = 1 \qquad \text{if } \alpha = \beta$$
$$= 0 \qquad \text{if } \alpha \neq \beta$$

(integrating with respect to Haar measure).

If we fix $\alpha \in \mathbb{Z}^d$ then applying this orthonormality relation to (12.2) gives:

$$\eta(s) = \sum_{k=1}^{\infty} \frac{1}{k} \sum_{\substack{\sigma^k x = x \\ \chi(g^k(x)) = \alpha}} hf^k(x)\, e^{-shf^k(x)}$$

$$= \int_{\mathbb{R}^d/\mathbb{Z}^d} \chi(-\alpha) \frac{L'(s,\chi)}{L(s,\chi)}\, d\chi \qquad (12.3)$$

(where $\eta(s)$ converges for $\mathcal{R}(s) > 1$).

By assumption (I) and the compactness of $\mathbb{R}^d/\mathbb{Z}^d$ we see that $\eta(s)$ is analytic in a neighbourhood of $\{1+it : t \neq 0\}$. If we choose a small neighbourhood U of χ_0 then

$$\int_{U^c} \chi(-\alpha) \frac{L'(s,\chi)}{L(s,\chi)}\, d\chi$$

is analytic in a neighbourhood of the half-plane $\mathcal{R}(s) \geq 1$, by Proposition 12.1 and assumption (I) again. It remains to analyse the contribution to $\eta(s)$ from $\int_U \chi(-\alpha) \frac{L'(s,\chi)}{L(s,\chi)} d\chi$, for small neighbourhoods U of χ_0 and s close to 1.

For $\chi \in U$, we can use the notation of (II) to write

$$\frac{L'(s,\chi)}{L(s,\chi)} = \frac{A}{(s-s(\chi))} + F(s,\chi)\,, \tag{12.4}$$

where $A \neq 0$ and $F(s,\chi)$ is analytic for $\mathcal{R}(s) \geq 1$ and the hypotheses on $\chi \mapsto s(\chi)$ in assumption (II) allows us to apply the Morse lemma: We may introduce coordinates $(\theta_1,\dots,\theta_d) \in U \subset \mathbb{R}^d/\mathbb{Z}^d$ satisfying $s(\chi) = 1 - \left(\theta_1^2 + \dots + \theta_d^2\right)$ (with $\chi = (\theta_1,\dots,\theta_d)$). In particular, we can write:

$$\frac{L'(s,\chi)}{L(s,\chi)} = \frac{B}{(s-1)+\left(\theta_1^2+\cdots+\theta_d^2\right)} + G(s,\chi) \tag{12.5}$$

where $B \neq 0$ and $G(s,\chi)$ is analytic for $\mathcal{R}(s) \geq 1$, by substituting into (12.4). To summarise, we have:

PROPOSITION 12.2. $\eta(s)$ *is analytic for* $\mathcal{R}(s) \geq 1$, *except for a singularity at* $s = 1$ *of the same order as*

$$A(s) = B \int_U \frac{B}{(s-1)+\left(\theta_1^2+\cdots+\theta_d^2\right)}\, d\theta_1 \dots d\theta_d\,.$$

We may restrict attention to neighbourhoods U of the form $U = \{(\theta_1,\dots,\theta_d) : \theta_1^2+\dots+\theta_d^2 \le \varepsilon^2\}$, for $\varepsilon > 0$ arbitrarily small and the evaluation of $A(s)$ becomes an exercise in integration.

LEMMA 12.3.

(i) *For* $d = 1$, $A(s) = C_1(s-1)^{-1/2}$

(ii) *For* $d = 2$, $A(s) = C_2 \log(s-1)$

(iii) *For* $d = 3$, $A(s) = C_3(s-1)^{1/2}$

(iv) *For* $d \ge 4$, $A(s) = C_d(s-1)^{(d-2)/2}$

(up to lower order terms) with C_d *constants independent of* U.

Now we have familiarised ourselves with the behaviour of $\eta(s)$ in the domain $\mathcal{R}(s) \ge 1$ we are in a position to study

$$\pi_\alpha(t) = \sum{}' 1$$

where the summation is restricted to those $x \in \mathrm{Fix}_k$ for which $e^{hf^k(x)} \le t$ and $\chi(g^k(x)) = \alpha$. The function

$$\psi_\alpha(t) = \sum{}' hf^k(x),$$

where the summation is similarly restricted, is a Stieltjes function for $\eta(s)$. In other words

$$\eta(s) = \int_1^\infty t^{-s}\, d\psi_\alpha(t)$$

for $\mathcal{R}(s) > 1$.

To use the method of Chapter 6 to derive asymptotic formulae for π_α we need to invoke a more general Tauberian theorem than the one due to Ikehara and Wiener. The following is due to Delange (cf. [26]).

THEOREM 12.4. *Let* $\psi(t) \geq 0, t \geq 0$ *be monotone non-decreasing with* $\eta(s) = \int_1^\infty t^{-s}\, d\psi(t)$ *analytic for* $\mathcal{R}(s) \geq 1$ *except for a singularity* $A(s)$ at $s = 1$ *(see Lemma 12.3).*

(i) *If* $A(s) = \dfrac{g_0(s)}{(s-1)^\beta} + g_1(s)$ *where* g_0, g_1 *are analytic for* $\mathcal{R}(s) \geq 1$

with $C = g_0(1) \neq 0$ *and* $\beta \in \mathbb{R} - \{-\mathbb{Z}^+\}$ *then* $\psi(t) \sim \dfrac{C.t}{\Gamma(\beta)\,(\log t)^{1-\beta}}$

(where $\Gamma(s)$ *denotes the gamma function).*

(ii) *If* $A(s) = C \log(s-1) + g(s)$ *where* $g_1(s)$ *is analytic for* $\mathcal{R}(s) \geq 1$

and $C \neq 0$ *then* $\psi(t) \sim \dfrac{C}{\log t}$.

By combining Lemma 12.3, Proposition 12.2 and Theorem 12.4 we can arrive at asymptotic estimates of $\psi_\alpha(t)$ in all cases except $d = 2\nu$, with $\nu \geq 2$. However, we can deal with these missing cases as follows:

We introduce the function:

$$\xi(s) = \sum_{k=1}^{\infty} \frac{1}{k} \sum_{\substack{\sigma^k x = x \\ \chi(g^k(x)) = \alpha}} n^\nu (hf^k(x))^{\nu+1} e^{-shf^k(x)}, \text{ for } \mathcal{R}(s) > 1,$$

then since we can write

$$\left(-\frac{d}{ds}\right)^\nu \frac{L'(S,\chi)}{L(S,\chi)} = \sum_{k=1}^{\infty} \frac{1}{k} \sum_{\sigma^k x = x} \chi(g^k(x))\,(hf^k(x)^\nu n^\nu e^{-shf^k(x)}$$

we have the identity

$$\xi(s) = \int_{\mathbb{R}^d/\mathbb{Z}^d} \chi(-\alpha)\left(-\frac{d}{ds}\right)^\nu \frac{L'(S,\chi)}{L(S,\chi)}\, d\chi$$

by the orthonormality relation.

As before we can choose coordinates $(\theta_1,\dots,\theta_d) \in U \subset \mathbb{R}^d/\mathbb{Z}^d$ in a neighbourhood U of χ_0 to arrive again at the identity (12.5). However, since $2\nu = d$ we can explicitly integrate

$$\int_U \frac{1}{(s-1)+\left(\theta_1^2+\cdots+\theta_d^2\right)}\, d\theta_1 \ldots d\theta_d = \frac{C}{S-1} + G(s)$$

where $U = \{(\theta_1,\ldots,\theta_d) : \theta_1^2+\cdots+\theta_d^2 \leq \varepsilon^2\}$, $C \neq 0$ and $G(s)$ is analytic on a neighbourhood of $\mathcal{R}(s) \geq 1$.

If we denote $\rho_\alpha(t) = \sum' n^\nu (hf^k(x))^{\nu+1}$ then the Ikehara-Wiener Tauberian theorem applied to

$$\xi(s) = \int_1^\infty t^{-s} d\rho_\alpha(t)$$

which is analytic for $\mathcal{R}(s) \geq 1$, except for a simple pole at $s = 1$, gives $\rho_\alpha(t) \sim Ct$. By a repeated application of the Abel summation formula we can deduce that $\psi_\alpha(t) \sim C_d t/(\log t)^\nu$.

We summarise our conclusions as follows:

PROPOSITION 12.5.

(i) *For* $d = 1$, $\psi_\alpha(t) \sim C_1 t/(\log t)^{1/2}$

(ii) *For* $d = 2$, $\psi_\alpha(t) \sim C_2\, t/\log t$.

(iii) *For* $d \geq 3$, $\psi_\alpha(t) \sim C_d \dfrac{t}{(\log t)^{d/2}}$.

It is more natural to obtain asymptotic estimates for $\pi_\alpha(t) = \sum_{\mathrm{Fix}_k}' 1$. By modifying the arguments at the end of Chapter 6 one derives the following:

THEOREM 12.6. *Under assumptions (I) and (II):*

(i) *For* $d = 1$, $\pi_\alpha(t) \sim C_1 \dfrac{t}{(\log t)^{3/2}}$

(ii) *For* $d = 2$, $\pi_\alpha(t) \sim C_2 \dfrac{t}{(\log t)^{2}}$

(iii) *For* $d \geq 3$, $\pi_\alpha(t) \sim C_d \dfrac{t}{(\log t)^{d/2+1}}$.

We return now to examples for which the hypotheses (I) and (II) are satisfied. We observe that when $\chi = \chi_0$ assumption (I) requires the flow to be weak-mixing (cf. Chapter 6).

Let $\varphi_t : M \to M$ be a geodesic flow for a compact surface S of negative curvature, with genus $g \geq 2$. The first homology group for S is $H_1(S,\mathbb{Z}) \cong \mathbb{Z}^{2g}$.

Let $T_1,\ldots,T_k$ be Markov sections for the flow. For any fixed point $s \in S$ we choose curves $\alpha_i : [0,1] \to S$, $1 \leq i \leq k$, with $\alpha_i(0) = p$ and $\alpha_i(1)$ in the projection of $T_i \subset M$ to S. Whenever $A(i,j) = 1$ we choose geodesic arcs γ_{ij} from the projection of T_i to the projection of T_j, and associate a closed curve $c_{ij} = \alpha_j^{-1} \circ \gamma_{ij} \circ \alpha_i$ based at p.

Let $\sigma : X_A \to X_A$ be the associated subshift of finite type, with suspended flow $\sigma_{f,t} : X_A^f \to X_A^f$. We define an extension $\tilde{\sigma} : X_A \times \mathbb{Z}^{2g} \to X_A \times \mathbb{Z}^{2g}$ by $\tilde{\sigma}(x,g) = (\sigma x, g+g(x))$, where $g(x) = [c_{x_0,x_1}] \in H_1(S,\mathbb{Z}) \cong \mathbb{Z}^{2g}$.

One can see that conditions (I) and (II) are satisfied as follows: For (I) we observe that $L(1+it,\chi) = \zeta(-(1+it)f + 2\pi ik)$ has a pole for $t \neq 0$ if and only if χ is a special character (in the sense of Chapter 8). However, this imposes constraints on the lengths of closed geodesics $\{f^n(x)\}$ which can easily be discounted. For (II) we begin by observing that $s(\chi_0) = 1$ and $\mathcal{R}s(\chi) < 1$ from condition (I) so that we immediately have $\nabla \mathcal{R}s(\chi)|_{\chi=\chi_0} = 0$. Rather than only showing $\nabla \mathrm{I}s(\chi)|_{\chi=\chi_0} = 0$ we can see that $\mathrm{I}s(\chi) \equiv 0$ for geodesic flows. This is a consequence of the existence of an involution $i : M \to M$ which reverses the direction of the geodesics. Since i carries closed orbits to closed orbits of the same period and $\chi \circ i = \bar{\chi}$ we deduce $\overline{L(s,\chi)} = L(s,\chi)$ and the claim follows.

Finally, we can write at the symbolic level $\det. \nabla^2 \mathcal{R}s(\chi)|_{\chi=\chi_0} = C\sigma^2(-hf) \leq 0$, where $C < 0$. However, this must be a strict inequality since otherwise f is cohomologous to a constant, which puts constraints on lengths of closed geodesics $\{f^n(x)\}$ and this again can be discounted. Applying part (iii) of Theorem 12.6 we deduce:

THEOREM 12.7. *Let* S *be a compact surface of negative curvature with genus* $g \geq 2$ *and closed geodesics* γ *of length* $\ell(\gamma)$ *representing an element* $[\gamma] \in H_1(S,\mathbb{Z})$. *For any element* $\alpha \in H_1(S,\mathbb{Z})$

$$\mathit{Card}\,\{\gamma : [\gamma] = \alpha, \ell(\gamma) \leq t\} \sim C\frac{e^{ht}}{t^{g+1}}$$

(for some constant $C > 0$).

Finally we want to compare Theorem 12.7 to a result of S.M. Rees on Fuchsian groups. (But see also the notes to this chapter.)

Let $\mathbb{D}^2$ denote the unit disc $\{ z \in \mathbb{C} \mid |z| < 1\}$ and let $ds^2 = \frac{1}{4}\,\frac{dx^2 + dy^2}{(1-(x^2+y^2))^2}$ denote the Poincaré metric on $\mathbb{D}^2$. We recall that a Fuchsian group is a finitely generated discrete subgroup $\Gamma \subset \mathrm{Isom}\,(\mathbb{D}^2, ds^2)$ of the isometries of $(\mathbb{D}^2, ds^2)$. The disc $\mathbb{D}^2$ has constant curvature $\kappa = -1$, relative to the metric ds^2 , so that the same is true of the quotient surface $S = \mathbb{D}^2 / \Gamma$, with respect to the induced metric.

Let $\Gamma_0 \lhd \Gamma$ be a normal subgroup then there is a corresponding covering surface $\mathbb{D}^2 \to \tilde{S} \to S$, where $\tilde{S} = \mathbb{D}^2/\Gamma_0$. We shall make the following two assumptions on Γ and Γ_0:

(i) $S = \mathbb{D}^2 / \Gamma$ is a compact surface ;

(ii) $\Gamma/\Gamma_0 \cong \mathbb{Z}^d$, for some $d \geq 1$.

(In particular, $\tilde{S}$ is non-compact, and does not even have finite area.)

Let $\varphi_t : M \to M$ denote the geodesic flow on the (compact) unit tangent bundle M of the surface S. This flow is hyperbolic (cf. Appendix IV). Let $T_1,\ldots,T_k$ denote a family of small Markov sections for $\varphi_t : M \to M$ and let $\tilde{\varphi}_t : \tilde{M} \to \tilde{M}$ denote the geodesic flow on the unit tangent bundle $\tilde{M}$ of the (non-compact) covering surface $\tilde{S}$. If $\pi : \tilde{M} \to M$ is the canonical projection (corresponding to the projection map $\tilde{S} \to S$) then we can lift the sections $T_1,\ldots,T_k$ to $\tilde{M}$. We denote this new family by $(T_{i,n})$, $1 \leq i \leq k$ and $n \in \mathbb{Z}^d$. This construction is similar to that in Chapter 8.

Next we want to introduce a complex function, associated to the groups Γ and Γ_0, called the *Poincaré series.* We fix a point in $\mathbb{D}^2$ (which, without loss of generality, we can take to be $0 \in \mathbb{D}^2$) and let $d(g0,0)$ denote the distance of the point 0 from its image $g0$ under the action of $g \in \mathrm{Isom}(\mathbb{D}^2)$ relative to the Poincaré metric. We define:

$$P(s,\Gamma) = \sum_{g \in \Gamma} e^{-sd(g0,0)}$$

$$P(s,\Gamma_0) = \sum_{g \in \Gamma_0} e^{-sd(g0,0)}$$

where $s \in \mathbb{C}$. (These series can readily be shown to converge in the half-plane $\mathcal{R}(s) > 1$.)

The group Γ_0 is said to be of *divergence type* if the limit of $P(s,\Gamma_0)$ does not exist as $s \searrow 1$, and of *convergence type* otherwise.

THEOREM 12.8. Γ_0 *is of divergence type if and only if* $\Gamma/\Gamma_0 \cong \mathbb{Z}^d$ *with* $d = 1$ or 2 *(cf. [76]).*

The behaviour of $P(s,\Gamma_0)$ is closely related to that of the function $\eta(s)$ (with the choice $\alpha = 0$). We can write

$$\tilde{\eta}(s) = \sum_{\tilde{\tau}} \lambda(\tilde{\tau})\, e^{-s\lambda(\tilde{\tau})}$$

where $\tilde{\tau}$ denotes a closed orbit for $\tilde{\varphi}$ of least period $\lambda(\tilde{\tau})$. In particular, $\tilde{\eta}(s)$ will have the same divergence and convergence properties at $s = 1$ as $\eta(s)$ (because of the standard argument about contributions from boundaries of sections and "auxiliary shifts") and we observe that $h = 1$ (cf. Appendix III). To relate $\tilde{\eta}(s)$ and $P(s,\Gamma_0)$ we observe the following:

(i) There is a natural bijection between closed orbits $\tilde{\tau}$ for $\tilde{\varphi}$ and conjugacy classes $[g]$ in Γ_0. Furthermore, there exists $c > 0$ such that $|\lambda(\tilde{\tau}) - d(0,g0)| < c$. (The bijection comes from the isomorphism $\pi_1(S) \cong \Gamma$ and $c \leq 2$ diameter(S).)

(ii) There exist constants $A,B > 0$ such that the number of elements $\eta(g)$ in the conjugacy class satisfies: $A \leq \frac{\eta(g)}{\lambda(\tilde{\tau})} \leq B$.

(For (ii) we note that by an observation due to Milnor the length $\lambda(\tilde{\tau})$ is related to the word length, for a fixed set of generators, which by a result of Nielson is related to $\eta(g)$.)

Clearly, (i) and (ii) show that $\hat{\eta}(s)$ and $P(s,\Gamma_0)$ have the same divergence properties as $\delta \searrow 1$. However, the behaviour of $\hat{\eta}(s)$ (or more precisely $\eta(s)$) is given by Lemma 12.3. In particular, we see that these functions diverge if and only if $d = 1,2$.

We can now consider the consequences of the subgroup Γ_0 being of divergence type or not. Let $\tilde{S} = \mathbb{D}^2/\Gamma$ then we say the associated geodesic flow $\tilde{\varphi}_t : \tilde{M} \to \tilde{M}$ (on the unit tangent bundle $\tilde{M}$ of $\tilde{S}$) is 'ergodic' if the only $\tilde{\varphi}_t$-invariant subsets of $\tilde{M}$ are those of zero measure (relative to the volume) or their complements are of zero measure.

These properties are related by the following: *The geodesic flow* $\tilde{\varphi}_t : \tilde{M} \to \tilde{M}$ *is ergodic if and only if* Γ_0 *is of divergence type* cf. [96].

In these cases there exists a unique probability measure m on the unit circle K called the *Patterson measure,* such that the action of the group $\Gamma_0 : K \to K$ satisfies $g^* m = |g'| m$, where $g \in \Gamma_0$ and the prime denotes differentiation [70].

In the case of Γ this measure can be constructed from the symbolic dynamics of Series, as follows: Series constructs an interval transformation $f : K \to K$ such that when f is restricted to certain arcs $I(y) \subset K$ it corresponds to the action on K of an associated generator $g \in \Gamma$ [89].

Let $L : C^0(K) \to C^0(K)$ denote the Ruelle-Perron-Frobenius operator defined by:

$$(Lw)(x) = \sum_{fy = x} \frac{w(y)}{|f'(y)|} \ , \ w \in C^0(K) .$$

We know the following standard result for expanding interval maps: *There exists a unique probability measure* m *such that* $L^* m = m$, *and* m *is equivalent to Lebesgue measure* (cf. [25]).

It is easy to see that $L^* m = m$ implies that $f^* m = |f'| m$. In view of the construction of $f : K \to K$ we see that m is precisely the Patterson measure.

To complete this chapter we shall state a more general version of Theorem 12.7 due to Katsuda and Sunada.

Let $\varphi_t : M \to M$ be a C^∞ Anosov flow which is weak-mixing. The *winding cycle* Φ for φ is a functional on 1-forms w such that $\Phi(w) = \int w(X)d\mu$, where X is the vector field generating the flow and μ is the measure of maximal entropy. The *covariance form* δ is defined by

$$\delta(w,w) = \lim_{t \to \infty} \frac{1}{t} \int \Big[\int_0^t w(X)(\varphi_u x)du - t\Phi(w)\Big]^2 d\mu(x) .$$

We can identify characters $\chi \in \mathrm{Hom}\,(H_1(M,\mathbb{Z}), \mathbb{C})$ with elements of $H^1(M,\mathbb{Z})$ (and 1-forms by deRham).

PROPOSITION 12.9. $\nabla^2{}_{\chi=\chi_0} \mathcal{R}s(\chi) = 4\pi^2\delta$ *and* $\nabla|S(\chi)|_{\chi=\chi_0} = \Phi$ *(cf. Katsuda and Sunada [46]).*

Let b be the rank of $H^* \subset H^1(M,\mathbb{Z})$ corresponding to the subgroup $H^* \subset H_1(M,\mathbb{Z})$ generated by closed orbits.

THEOREM 12.10. *If* Φ *vanishes on* H^* *then*

$$\mathit{Card}\,\{\tau : [\tau] = \alpha\, ,\, \lambda(\tau) \le t\} \sim C \frac{e^{ht}}{t^{b/2+1}} \quad \textit{for each } \alpha \in H$$

(for some constant $C > 0$*).*

Notes

The results in this chapter on non-compact extensions are intended to complement the results in Chapter 8 on compact extensions. Most of the material we present is derived from work of Katsuda and Sunada [45], [46]. These authors give a fairly comprehensive analysis, part of which we summarise at the end of the chapter. Preliminary results for geodesic flows appeared in [47].

The result of Mary Rees on divergence type appears in [76]. (The referee has kindly pointed out that Guivarc'h also obtained this result. See [35*].) The Patterson measure was introduced in [70] and the connection between divergence type and ergodicity is described in [96] and [97]. The role of the Ruelle operator in describing the Patterson measure is explained in the last section of [88].

APPENDIX I

THE IKEHARA-WIENER TAUBERIAN THEOREM

We need some standard facts from the theory of the Fourier integral. (See, for example [48].)

For $f \in L^1(\mathbb{R})$ the *Fourier transform* is defined by

$$\hat{f}(x) = \int_{-\infty}^{\infty} g(y)\, e^{-iyx} dy$$

$$\hat{f}(0) = \int_{-\infty}^{\infty} g(y)\, dy .$$

(I.1) *The Fourier transform* $\hat{f}$ *is uniformly continuous and* $\lim_{|x| \to \infty} \hat{f}(x) = 0$ *(the Riemann-Lebesgue lemma).*

(I.2) *When* $f, \hat{f} \in L^1(\mathbb{R})$ *then*

$$\frac{1}{2\pi} \hat{\hat{f}}(-x) = f(x) .$$

If f is integrable and g is bounded or integrable, or if f,g are measurable and non-negative their *convolution* $f*g$ is defined by

$$f*g(x) = \int_{-\infty}^{\infty} f(x-y)\, g(y)\, dy\,.$$

(I.3) *When* $f, g \in L^1(\mathbb{R})$ *then*

$$\widehat{f*g} = \hat{f}.\hat{g}\,.$$

We shall use the *Fejer kernel*

$$K_N(x) = \frac{1}{2\pi N}\left(\frac{\sin(Nx/2)}{\sin(x/2)}\right)^2 \in L^1(\mathbb{R}),\ N = 1,2,\ldots$$

whose Fourier transform is

(I.4) $$\begin{aligned} \hat{K}_N(t) &= \left(1 - \frac{|t|}{N}\right) && \textit{when } |t| \le N \\ &= 0 && \textit{when } |t| \ge N\,. \end{aligned}$$

(I.5) *If* $f \in L^1(\mathbb{R})$ *then* $f*K_N \to f$ *in the* $L^1(\mathbb{R})$ *norm as* $N \to \infty$.

(I.6) LEMMA

If $\varphi \in L^\infty(\mathbb{R})$ *and* $\varphi(x) \to A$ *as* $x \to \infty$ *then* $g*\varphi(x) \to A\hat{g}(0)$ *as* $x \to \infty$

for all $g \in L^1(\mathbb{R})$.

PROOF. Evidently

$$|g*\varphi(x) - A\hat{g}(0)| = \left| \int_{-\infty}^{\infty} g(x-y)\,\varphi(y)dy - \int_{-\infty}^{\infty} g(x-y)\,A dy \right|$$

$$\le \left| \int_{-\infty}^{x/2} g(x-y)\,(\varphi(y)-A)dy \right| + \int_{x/2}^{\infty} |g(x-y)|\,|\varphi(y)-A|\,dy$$

$$\le \|\varphi-A\|_{\infty} \int_{-\infty}^{x/2} |g(x-y)|\,dy + \varepsilon \int_{x/2}^{\infty} |g(x-y)|\,dy \quad \text{(if } x \text{ is large)}$$

$$\le \|\varphi-A\|_{\infty} \int_{x/2}^{\infty} |g(u)|\,du + \varepsilon\,\|g\|_1$$

and this latter expression converges to $\varepsilon\,\|g\|_1$ as $x \to \infty$.

(I.7) WIENER'S TAUBERIAN THEOREM (Weak version) (cf. [103].)

If $\varphi \in L^{\infty}(\mathbb{R})$ *and* $K_N*\varphi(x) \to A$ *as* $x \to \infty$ *for each* $N = 1,2,\ldots,$ *then* $g*\varphi(x) \to A\hat{g}(0)$ *as* $x \to \infty$ *for all* $g \in L^1(\mathbb{R})$.

PROOF. We have the following inequalities,

$$|g*\varphi(x) - A\hat{g}(0)| \le |g*\varphi(x) - K_N*g*\varphi(x)| + |K_N*g*\varphi(x) - A\hat{g}(0)|$$
$$\le \|g - K_N*g\|_1\,\|\varphi\|_{\infty} + |K_N*g*\varphi(x) - A\hat{g}(0)|.$$

If N is large then $\|g - K_N * g\|_1$ is small and by the lemma, for fixed N,

$$|K_N * g * \varphi(x) - A\hat{g}(0)| \to 0 \text{ as } x \to \infty .$$

Hence $|g * \varphi(x) - A\hat{g}(0)| \to 0$ as $x \to \infty$.

(I.8) THE IKEHARA-WIENER TAUBERIAN THEOREM (See, for example [103].)

Let $\alpha(x)$ *be a monotonic non-decreasing and continuous from above with* $\alpha(1) = 0$. *Suppose*

$$\int_1^\infty x^{-s} d\alpha(x) = \frac{A}{s-1} + \varphi(s)$$

for $\mathcal{R}(s) > 1$, *where the integral is absolutely convergent and where the continuous function* $\varphi(s)$ *converges uniformly on bounded intervals as* $\mathcal{R}(s) \searrow 1$ *so that* $\varphi(1+it)$ *is continuous. Then* $\alpha(x) \sim Ax$ *as* $x \to \infty$.

PROOF. Integration by parts yields $\int_1^\infty x^{-s} d\alpha(x) = \left[x^{-s}\alpha(x)\right]_1^\infty + s\int_1^\infty x^{-s-1}\alpha(x)dx$ and the hypotheses ensure that $\left[x^{-s}\alpha(x)\right]_1^\infty = 0$, when $\mathcal{R}(s) > 1$. Hence

$$\int_1^{\infty} x^{-s} d\alpha(x) = s \int_0^{\infty} e^{-(s-1)x} \alpha(e^x) e^{-x} dx \,.$$

In particular

$$\frac{1}{s-1} \int_1^{\infty} x^{-s} dx = s \int_0^{\infty} e^{-(s-1)x} dx$$

so that

$$\frac{\varphi(s)}{s} = \int_0^{\infty} e^{-(s-1)x}(\alpha(e^x)e^{-x} - A)dx.$$

With $\rho(s) = \frac{\varphi(s)}{s}$ we have

$$\rho(1+\varepsilon+it) = \int_{-\infty}^{\infty} e^{-itx} e^{-\varepsilon x} (\alpha(e^x)e^{-x} - A) \chi_{[0,\infty)}(x) \, dx$$

and defining $\psi_\varepsilon(x) = e^{-\varepsilon x} (\alpha(e^x)e^{-x} - A) \chi_{[0,\infty)}(x)$, $\rho_\varepsilon(t) = \rho(1+\varepsilon+it)$ and $A_\varepsilon(x) =$

$Ae^{-\varepsilon x} \chi_{[0,\infty)}(x)$ for $\varepsilon \geq 0$, we see that for $\varepsilon > 0$

$$\hat{\psi}_\varepsilon(t) = \hat{A}_\varepsilon(t) + \rho_\varepsilon(t)\,.$$

(Note that $\psi_\varepsilon, A_\varepsilon \in L^1(\mathbb{R})$ when $\varepsilon > 0$.)

A_0 is bounded and we shall now prove that ψ_0 is bounded.

Notice that for fixed N, $\hat{\psi}_\varepsilon.\hat{K}_N = \hat{A}_\varepsilon.\hat{K}_N + \rho_\varepsilon.\hat{K}_N$

so taking inverse transforms we have

$$\psi_\varepsilon * K_N(x) = A_\varepsilon * K_N(x) + \frac{1}{2\pi}\int_{-N}^{N} \rho(1+\varepsilon+it)\left(1 - \frac{|t|}{N}\right) e^{ixt}\,dt$$

and by the monotone convergence theorem

$$\psi_0 * K_N(x) = A_0 * K_N(x) + \frac{1}{2\pi}\int_{-N}^{N} \rho(1+it)\left(1 - \frac{|t|}{N}\right) e^{ixt}\,dt\,.$$

The limit of this last integral is zero, as $x \to \infty$, by the Riemann-Lebesgue lemma. Hence

$$\lim_{x\to\infty} \psi_0 * K_N(x) = \lim_{x\to\infty} \int_0^\infty K_N(x-y) A\,dy$$

$$= \lim_{x\to\infty} \int_{-\infty}^{x} K_N(u) A \, du$$

$$= A .$$

Let $a > 0$, then

$$\overline{\lim_{x\to\infty}} \; \psi_0(x) e^{-a} \int_{x}^{x+a} K_N(x-y)\, dy$$

$$\le \overline{\lim_{x\to\infty}} \int_{x}^{x+a} K_N(x-y)\, \psi_0(y)\, dy$$

$$\le \overline{\lim_{x\to\infty}} \; \psi_0 * K_N(x) = A .$$

(Here we have used the fact that $\psi_0(y)e^y$ is increasing with y.)

We see therefore that $\overline{\lim\limits_{x\to\infty}} \; \psi_0(x)$ is finite so that $\alpha(e^x)e^{-x}$ is bounded.

From the above we have $\psi_0 - A_0$ is bounded and

$$(\psi_0 - A_0) * K_N(x) = \frac{1}{2\pi} \int_{-N}^{N} \rho(1+it) \left(1 - \frac{|t|}{N} \right) e^{ixt}\, dt$$

which tends to zero as $x \to \infty$ by the Riemann-Lebesgue lemma. Hence, by the Wiener Tauberian theorem $\psi_0 * f(x) \to A\hat{f}(0)$ as $x \to \infty$ when $f \in L^1(\mathbb{R})$.

Now let $f = f_1$ and $f = f_2$ in turn where f_1, f_2 are non-negative with supports in $[-\varepsilon, 0]$ and $[0, \varepsilon]$ respectively and $\int_{-\infty}^{\infty} f_1(x)dx = \int_{-\infty}^{\infty} f_2(x)dx = 1$. We use the fact that $e^y \psi_0(y)$ is increasing. Clearly $\psi_0(y) \le e^{\varepsilon}\psi_0(x)$ if $x-\varepsilon \le y \le x$ and $e^{-\varepsilon}\psi_0(x) \le \psi_0(y)$ if $x \le y \le x+\varepsilon$. Hence $e^{-\varepsilon} f_2 * \psi_0(x) \le \psi_0(x) \le e^{\varepsilon} f_1 * \psi_0(x)$. Therefore

$$e^{-\varepsilon} A\hat{f}_2(0) \le \varliminf_{x \to \infty} \psi_0(x) \le \varlimsup_{x \to \infty} \psi_0(x) \le e^{\varepsilon} A\hat{f}_1(0).$$

Thus $\lim_{x \to \infty} \psi_0(x) = A$. In other words $\alpha(e^x) \sim Ae^x$ as $x \to \infty$.

APPENDIX II

UNITARY COCYCLES

This appendix is devoted to the proof of Proposition 8.2 which says, in effect, that a continuous unitary matrix valued function defined on X is cohomologous to another such function which depends only on future coordinates if the initial function has n'th variations decreasing to zero at an exponential rate.

To be precise let (X,σ) be a shift of finite type and let $U(d)$ denote the group of $d \times d$ unitary matrices equipped with the usual topology. If $F : X \to U(d)$ is continuous we define its n'th variation by

$$\mathrm{var}_n F = \sup\{|F(x) - F(y)| : x_i = y_i, |i| < n\}$$

where $| \, |$ denotes the Euclidean norm on matrices. For $0 < \theta < 1$ let

$$U(\theta,d) = \{F : X \to U(d) : F \text{ is continuous and for all } n \geq 0$$
$$\mathrm{var}_n F \leq K\theta^n \text{ for some constant } K\}.$$

Here we prove the following analogue of Proposition 1.2:

(II.1) THEOREM. *Let* $F \in U(\theta,d)$ *then there exists* $G,F' \in U(\theta^{\frac{1}{2}},d)$ *such that* $F'(x) = F'(y)$ *whenever* $x_i = y_i$ *for all* $i \geq 0$ *and such that*

$$F'(x) = G(\sigma x)^{-1}F(x)G(x)\,.$$

We shall first need the following elementary

(II.2) LEMMA. *Let* $U_0,\dots,U_n$, $u_0,\dots,u_n$, $V_0\cdots V_n$, $v_0\cdots v_n$ *be unitary matrices of the same dimension. Then*

$$|U_0^{-1}\cdots U_n^{-1}u_n\dots u_0 - V_0^{-1}\cdots V_n^{-1}v_n\cdots v_0|$$

$$\leq \sum_{i=0}^{k}|U_i-V_i| + \sum_{i=0}^{k}|u_i-v_i| + \sum_{i=k+1}^{n}|U_i-u_i| + \sum_{i=k+1}^{n}|V_i-v_i|\,.$$

PROOF. To see this one uses the fact that the norm of a unitary matrix is 1. For example

$$|U_0^{-1}\cdots U_n^{-1}u_n\cdots u_0 - V_0^{-1}\cdots V_n^{-1}v_n\cdots v_0|$$

$$\leq |V_0U_0^{-1}U_1^{-1}\cdots U_n^{-1}u_n\cdots u_1u_0v_0^{-1} - U_1^{-1}\cdots U_n^{-1}u_n\dots u_1|$$

$$+\,|U_1^{-1}\cdots U_n^{-1}u_n\cdots u_1 - V_1^{-1}\cdots V_n^{-1}v_n\dots v_1|$$

$$\leq |V_0U_0^{-1}-I| + |u_0v_0^{-1} - I| + |U_1^{-1}\cdots U_n^{-1}u_n\dots u_1 - V_1^{-1}\cdots V_n^{-1}v_n\dots v_1|\,.$$

In this way we see that the initial quantity is dominated by

$$\sum_{i=0}^{k} |V_i - U_i| + \sum_{i=0}^{k} |v_i - u_i| + |U_{k+1}^{-1} \cdots U_n^{-1} u_n \cdots u_{k+1} - V_{k+1}^{-1} \cdots V_n^{-1} v_n \cdots v_{k+1}| .$$

The last term in this latter quantity, however, is dominated by

$$\sum_{i=k+1}^{n} |U_i - u_i| + \sum_{i=k+1}^{n} |V_i - v_i|$$

as can be seen by the introduction of the identity matrix, using the triangle inequality and performing elementary manipulations.

As in the proof of Proposition 1.2, for each state i let $\left\{ j_n^i \right\}_{-\infty}^{0}$ be an allowable sequence such that $j_0^i = i$ and define, for each $x \in X$, $\varphi(x) \in X$ by $\varphi(x)_n = x_n$ for $n \geq 0$ and $\varphi(x)_n = j_n^i \,(i = x_0)$ for $n < 0$. Define the function $G_n : X \to U(d)$ by

$$G_n(x) = F(x)^{-1} \cdots F(\sigma^n x)^{-1} F(\sigma^n \varphi x) \cdots F(\varphi x) .$$

Consider two points $x,y \in X$ such that $d(x,y) \leq \theta^{2k}$ i.e. $x_i = y_i$ whenever $|i| \leq 2k$. Clearly

$$|F(\sigma^n x) - F(\sigma^n y)| \leq \|F\|_\theta \theta^{2k-n} \text{ for } n < 2k$$
$$|F(\sigma^n \varphi x) - F(\sigma^n \varphi y)| \leq \|F\|_\theta \theta^{2k-n} \text{ for } n < 2k$$

and $\quad |F(\sigma^n z) - F(\sigma^n \varphi z)| \quad \leq \|F\|_\theta \theta^n$ for *all* n and z .

Writing $U_n = F(\sigma^n x)$, $u_n = F(\sigma^n \varphi x)$, $V_n = F(\sigma^n y)$, $v_n = F(\sigma^n \varphi y)$ and applying lemma II.2 we see that

$$\begin{aligned} |G_n(x) - G_n(y)| &\leq 2\|F\|_\theta(\theta^{2k}+\cdots+\theta^k) + 2\|F\|_\theta(\theta^{k+1} + \theta^{k+2}+\cdots) \\ &\leq 2\|F\|_\theta \frac{\theta^k}{1-\theta} + 2\|F\|_\theta \frac{\theta^{k+1}}{1-\theta} \\ &\leq 4\|F\|_\theta \frac{\theta^k}{1-\theta} . \end{aligned}$$

From this inequality it is clear that $\{G_n\}$ is a uniformly equicontinuous sequence of functions which converges to a continuous function G. Moreover G *also* satisfies the inequality

$$|G(x) - G(y)| \leq 4\|F\|_\theta \frac{\theta^k}{1-\theta}$$

for $d(x,y) \leq \theta^{2k}$, so that $G \in U(\theta^{\frac{1}{2}},d)$.

Now define

$$\begin{aligned} &F'_n(x) = G_n(\sigma x)^{-1}F(x)G_n(x) \\ &= [F(\sigma x)^{-1}\cdots F(\sigma^{n+1})^{-1} F(\sigma^n\varphi\sigma x)\cdots F(\varphi\sigma x)]^{-1} F(x)F(x)^{-1}\cdots F(\sigma^n x)^{-1}F(\sigma^n\varphi x)\cdots F(\varphi x) \\ &= F(\varphi\sigma x)^{-1}\cdots F(\sigma^n\varphi\sigma x)^{-1}F(\sigma^{n+1}x)\cdots F(\sigma x)F(x)F(x)^{-1}\cdots F(\sigma^n x)^{-1}F(\sigma^n\varphi x)\cdots F(\varphi x) \end{aligned}$$

$$= F(\varphi\sigma x)^{-1} \cdots F(\sigma^n \varphi\sigma x)^{-1} F(\sigma^{n+1}x) F(\sigma^n \varphi x) \cdots F(\varphi x)\ .$$

Apart from the central term in this last expression, namely

$$F(\sigma^n \varphi\sigma x)^{-1} F(\sigma^{n+1}x)$$

we have a form which depends only on the future coordinates of x. It is a simple matter to show that the exceptional central term is increasingly negligible as $n \to \infty$, so we see that F'_n converges to a function F' such that

$$F' = (G \circ \sigma)^{-1}.F.G$$

and such that $F'(x) = F'(y)$ whenever $x_i = y_i$ for all $i \geq 0$. The proof of Theorem II.1 is therefore complete.

APPENDIX III

HYPERBOLIC DYNAMICS, MARKOV PARTITIONS AND ZETA FUNCTIONS

In this appendix we shall collect together several results on hyperbolic systems and symbolic dynamics which were needed in the main text. We shall state the principle results and refer the reader to the appropriate sources. In most cases we shall attempt to present a sketch of the proofs which convey the main ideas, without becoming too involved in technical details.

§1. Markov partitions and symbolic dynamics. An important feature of hyperbolic systems is that they can be effectively modelled by symbolic dynamics, i.e. subshifts of finite type for hyperbolic diffeomorphisms and suspended flows for hyperbolic flows. We have been principally concerned with hyperbolic flows. However, the construction of the symbolic dynamics for these flows is more complex than that for hyperbolic diffeomorphisms. For this reason we shall begin by describing the constructions for hyperbolic diffeomorphisms as a precursor to the flow case.

§1.1. Axiom A diffeomorphisms. Our account is a summary of Bowen's work in [10]. Let M be a C^∞ compact Riemannian manifold and let $f : M \to M$ be a C^1 diffeomorphism.

DEFINITION. We call a point $x \in M$ *wandering* if there exists an open neighbourhood U of x such that $f^n U \cap U = \emptyset$ for $n > 1$. The *non-wandering set* Ω is the complement of the union of the wandering points and is closed and f invariant.

The diffeomorphism f satisfies *Axiom A* if:

(a) Ω is *hyperbolic,* i.e. there exists a continuous splitting $T_\Omega M = E^u \oplus E^s$ into a Whitney sum of Df-invariant sub-bundles and there exist $C > 0$, $0 < \lambda < 1$ such that

$\|Df^n(v)\| \leq C\lambda^n \|v\|$ for $v \in E^s$, $n \geq 0$ and $\|Df^{-n}(v)\| \leq C\lambda^n \|v\|$ for $v \in E^u$, $n \geq 0$.

(b) The periodic points of f are dense in Ω.

(III.1) PROPOSITION. (Smale spectral decomposition).

The non-wandering set Ω *has a decomposition* $\Omega = \bigcup_{i=1}^{n} \Omega_i$ *where the* Ω_i *are closed,* f-invariant *disjoint sets and* $f|\Omega_i$ *is transitive.*

One can also decompose each Ω_i as $\Omega_i = \bigcup_{j=1}^{n(i)} \Omega_i^j$ where the Ω_i^j are closed and disjoint with $f(\Omega_i^j) = \Omega_i^{j+1}$ $(1 \leq j \leq n(i)-1)$ and $f(\Omega_i^{n(i)}) = \Omega_i^1$. Moreover $f^{n(i)} : \Omega_i^j \to \Omega_i^j$ is topologically mixing.

The above proposition allows us to restrict our attention to $f : \Omega_i \to \Omega_i$.

More generally, one can consider any diffeomorphism $f : M \to M$ with a closed invariant set $\Lambda \subset M$ such that:

(a) $f : \Lambda \to \Lambda$ is hyperbolic;

(b) $f : \Lambda \to \Lambda$ is transitive;

(c) the periodic points of $f|_\Lambda$ are dense in Λ;

(d) there exists a neighbourhood $U \supset \Lambda$ with $\Lambda = \bigcap_{n \equiv -\infty}^{\infty} f^n(U)$.

We call $f : \Lambda \to \Lambda$ a *hyperbolic diffeomorphism,* and our principle examples are $f : \Omega_i \to \Omega_i$. Henceforth we shall assume that Λ is not a single closed orbit.

STABLE AND UNSTABLE MANIFOLDS AND LOCAL PRODUCT STRUCTURE. The splitting of $T_\Lambda M$ is reflected in the existence of certain submanifolds in M itself which exhibit expansion and contraction under the action of f. For $\varepsilon > 0$ we define the (local) *stable manifold* for $x \in \Lambda$ by

$$W^s_\varepsilon(x) = \{y \in M : d(f^n x, f^n y) \le \varepsilon \text{ for all } n \ge 0\}$$

and the (local) *unstable manifold* for $x \in \Lambda$ by

$$W^u_\varepsilon(x) = \{y \in M : d(f^{-n} x, f^{-n} y) \le \varepsilon \text{ for all } n \ge 0\}.$$

Hirsch, Pugh and Shub [39] showed that for $\varepsilon > 0$ sufficiently small $W^s_\varepsilon(x)$ and $W^u_\varepsilon(x)$ are C^1 embedded discs with $T_x W^s_\varepsilon(x) = E^s_x$ and $T_x W^u_\varepsilon(x) = E^u_x$.

There exists $\delta = \delta(\varepsilon) > 0$ such that whenever $d(x,y) < \delta$, $x,y \in \Lambda$ then $W^s_\varepsilon(x) \cap W^u_\varepsilon(y) \neq \emptyset$. Furthermore, this interesection is a single point of Λ, which we denote by $[x,y]$. This property defines a so called *local product structure,* and correspondingly a local map $(x,y) \to [x,y]$.

MARKOV PARTITIONS. The underlying idea is to cover Λ by a finite number of closed sets, numbered from 1 to k, say. A point $x \in \Lambda$ with orbit $\ldots, f^{-2}x, f^{-1}x, x, fx, f^2x, \ldots$ will give rise to a sequence from $\prod_{-\infty}^{+\infty} \{1,\ldots,k\}$, where the ith term of the sequence will correspond to the index of the set containing f^ix. We want to choose sets which give rise to particularly simple sequences which accurately model the diffeomorphism $f : \Lambda \to \Lambda$.

DEFINITION. A set $R \subset \Lambda$ is called *a rectangle* if whenever $x,y \in R$ then $[x,y] \in R$ and *proper* if $R = (\overline{\text{int } R})$.

We want to construct a 'partition' of proper rectangles $\{R_i\}_{i=1}^k$ for Λ, in the sense that $\Lambda = \bigcup_{i=1}^k R_i$ and $\text{int } R_i \cap \text{int } R_j = \emptyset$ for $i \neq j$. In order that the sequences in $\prod_{-\infty}^{+\infty} \{1,\ldots,k\}$ corresponding to f-orbits in Λ should take a particularly simple form we require an additional condition:

DEFINITION. The proper rectangles $\{R_i\}_{i=1}^{k}$ form a *Markov partition* for $f : \Lambda \to \Lambda$ if

(a) For $x \in \operatorname{int} R_i$ with $fx \in \operatorname{int} R_j$ then $f(W^s(x,R_i)) \subset W^s(fx,R_j)$

and (b) For $x \in \operatorname{int} R_i$ with $f^{-1}x \in \operatorname{int} R_j$ then $f^{-1}(W^u(x,R_i)) \subset W^u(f^{-1}x,R_j)$

where we write $W^s(x,R_i) = W^s_\varepsilon(x) \cap R_i$, $W^u(x,R_i) = W^u_\varepsilon(x) \cap R_i$.

(Note. We shall always concern ourselves with the case $\text{diam.}(R_i) << \varepsilon << \text{diam.}(\Omega)$.)

(III.2) PROPOSITION. *For a hyperbolic diffeomorphism* $f : M \to M$ *there exists an arbitrarily small Markov partition for* Λ.

PROOF. (cf. [10]) We can assume $\lambda < \frac{1}{2}$, otherwise we replace f by f^N with $\lambda^N < \frac{1}{2}$. Let $\eta > 0$ be a constant, to be specified later. As a first approximation let $\left\{R_i^0\right\}$ be a finite cover for Ω where $R_i^0 = \left[S_i^0 , U_i^0\right]$, with S_i^0 , U_i^0 being closed subsets of (local) stable and unstable manifolds, respectively, with $\text{diam}\left(S_i^0 \cup U_i^0\right) \leq \eta$.

Generally, $\left\{R_i^0\right\}$ will not be a Markov partition. For a second approximation (incorporating first order corrections) we choose

$$S_i^1 = S_i^0 \cup \left\{\left[S_i^0 , fS_j^0\right] : \text{int } R_i^0 \cap f(\text{int } R_j^0) \neq \emptyset\right\} \text{ and}$$

$$U_i^1 = U_i^0 \cup \left\{\left[f^{-1}U_j^0 , U_i^0\right] : \text{int } R_i^0 \cap f^{-1}(\text{int } R_j^0) \neq \emptyset\right\}$$

and set $R_i^1 = \left[S_i^1 , U_i^1\right]$. Here $S_i^1 \supset S_i^0 , U_i^1 \supset U_i^0$ and $[\, ,\,]$ is always well-defined if η is small enough.

Inductively we define, for each $k \geq 1$:

$$S_i^k = S_i^{k-1} \cup \left\{\left[S_i^{k-1} , fS_j^{k-1}\right] : \text{int } R_i^0 \cap f(\text{int } R_j^0) \neq \emptyset\right\} \text{ and}$$

$$U_i^k = U_i^{k-1} \cup \left\{\left[f^{-1}U_j^{k-1} , U_i^{k-1}\right] : \text{int } R_i^0 \cap f^{-1}(\text{int } R_j^0) \neq \emptyset\right\} \text{ and } R_i^k = \left[S_i^k , U_i^k\right].$$

Hence $S_i^k \supset S_i^{k-1} , U_i^k \supset U_i^{k-1}$ and $[\, ,\,]$ is well-defined, again, if η is small enough. By construction: diam S_i^k , diam $U_i^k \leq \eta.K\ (1+C(2\lambda)+C(2\lambda)^2+\cdots)$ which can be made arbitrarily small by our choice of η. (Here $K > 0$ is a constant independent of k.)

Therefore, we can take $S_i^\infty = \bigcup_{k=0}^{\infty} S_i^k, U_i^\infty = \bigcup_{k=0}^{\infty} U_i^k$ and set $R'_i = [S_i^\infty, U_i^\infty]$. The rectangles R'_i satisfy a 'Markovian' condition, but do not necessarily have disjoint interiors nor are they necessarily proper. To overcome the first problem we take suitable intersections of overlapping rectangles from R'_i to arrive at a family $\{R''_i\}$. Furthermore, since the intersection is relative to the interiors of rectangles the family $\{R''_i\}$ is the desired proper Markov partition.

The above 'proof', due to Bowen [10], is a generalisation of Sinai's proof for Anosov diffeomorphisms [92]. There is an alternative proof, also due to Bowen, in [16] (cf. also [90]).

Let $\{R_i\}_{i=1}^{k}$ be a 'small' Markov partition. We define a $0-1$ $k \times k$ matrix A by:

$$A(i,j) = \begin{cases} 1 & \text{if } f(\operatorname{int} R_i) \cap (\operatorname{int} R_j) \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

and we let $\sigma : X_A \to X_A$ be the associated subshift of finite type.

(III.3) PROPOSITION. *The map* $\pi : X_A \to \Lambda$ *defined by* $\pi(x) = \bigcap_{n=-\infty}^{\infty} f^{-n} R_{x_n}$ *is well-defined.*

PROOF. Let $B_m(x) = \bigcap_{n=-m}^{m} f^{-n}R_{x_n}$ for $m \geq 0$. Clearly $B_m \supset B_p$ for $p \geq m$ and $\operatorname{diam} B_m \leq K\lambda^m$ for some constant $K > 0$. Therefore $\pi(x)$ is at *most* a single point. We show $\pi(x)$ is non-empty inductively: Assume for $x \in X_A$ that $B_m(x) \neq \emptyset$. Choose $w \in f^{-1}B_m(\sigma x)$, $z \in fB_m(\sigma^{-1}x)$, then $[w,z] \in B_{m+1}(x)$. In particular, $B_{m+1}(x) \neq \emptyset$. Proceeding inductively, $B_p(x) \neq \emptyset$ for all $p \geq 0$ and therefore $\pi(x) \neq \emptyset$.

The effectiveness of $\sigma : X_A \to X_A$ in modelling $f : \Lambda \to \Lambda$ is summarised in the following theorem:

(III.3) THEOREM (Bowen).

(i) *π is Hölder continuous and surjective.*

(ii) *π is one-one on a set of full measure (for any ergodic measure of full support) and on a dense residual set.*

(iii) *π is bounded-one.*

(iv) $f\pi = \pi\sigma$.

PROOF.

(i) Since $\operatorname{diam} B_m \leq K\lambda^m$ we see that π is Hölder continuous with Hölder exponent $\alpha = (\log \lambda / \log \theta)$ where $0 < \theta < 1$ is chosen in defining the metric on X_A. Since X_A, and hence $\pi(X_A)$, is compact and $\pi(X_A)$ is dense we see that π is surjective.

(ii) π only fails to be one-one when $f^n x \in \partial R$, for some n where $\partial R = \bigcup_i \partial R_i$. Since ∂R is closed and nowhere dense, $\bigcup_{n=-\infty}^{\infty} f^n \partial R$ has a dense residual complement, by the Baire Category theorem.

If $\partial^s R_i = \{x \in \partial R_i : W^u_\varepsilon(x,R_i) \cap \operatorname{int} R_i \neq \emptyset\}$, $\partial^u R_i = \{x \in \partial R_i : W^s_\varepsilon(x,R_i) \cap \operatorname{int} R_i \neq \emptyset\}$ then $\partial R_i = \partial^s R_i \cup \partial^u R_i$. If $\partial^s R = \bigcup_{i=1}^{k} \partial^s R_i$, $\partial^u R = \bigcup_{i=1}^{k} \partial^u R_i$ then $f\partial^s R \subseteq \partial^s R$ and $f^{-1}\partial^u R \subseteq \partial^u R$. Since $\partial^s R, \partial^u R$ are nowhere dense they have zero measure for any ergodic measure of full support.

(iii) We shall show $\operatorname{Card} \pi^{-1}(x) \leq k^2$. Otherwise, choose $\{x^{(i)}\}_{i=1}^{k^2+1} \in \pi^{-1}(x)$ to be distinct. Next choose $N > 0$ sufficiently large that $\left\{\left(x^{(i)}_{-N},\ldots,x^{(i)}_N\right)\right\}_{i=1}^{k^2+1}$ are distinct $(2N+1)$-tuples. But (by the pigeon-hole principle) there must be $1 \leq i < j \leq k^2+1$ with $x^{(i)}_{-N} = x^{(j)}_{-N}$ and $x^{(i)}_N = x^{(j)}_N$. We can choose $y^{(i)} \in \bigcap_{\alpha=-N}^{N} f^{-\alpha} \operatorname{int} R_{x^{(i)}_\alpha}$, $y^{(j)} \in \bigcap_{\alpha=-N}^{N} f^{-\alpha} \operatorname{int} R_{x^{(j)}_\alpha}$ by the Markov condition. Then $[f^\alpha y^{(j)}, f^\alpha y^{(i)}] \in f^{\alpha+N} W^s_\varepsilon(f^{-N} y^{(j)}, R_{x^{(j)}_{-N}}) \subset R_{x^{(j)}_\alpha}$ and $[f^\alpha y^{(j)}, f^\alpha y^{(i)}] \in f^{\alpha-N} W^u_\varepsilon(f^N y^{(i)}, R_{x^{(j)}_N}) \subset R_{x^{(i)}_\alpha}$ for $-N \leq \alpha \leq N$. Thus $\left(x^{(i)}_{-N},\ldots,x^{(i)}_N\right) = \left(x^{(j)}_{-N},\ldots,x^{(j)}_N\right)$, contradicting our hypothesis.

(iv) This follows directly by construction since $f\pi(x) = f\left(\bigcap_{n=-\infty}^{\infty} f^{-n}R_{x_n}\right) = \bigcap_{n=-\infty}^{\infty} f^{-n}R_{x_{n+1}} = \pi(\sigma x)$.

It is easy to see that since $f : \Lambda \to \Lambda$ is transitive then so is $\sigma : X \to X$. If $A_1,\ldots,A_n$ are the irreducible component matrices of A then for each $1 \le i \le n$, we can write $A_i = A_i^1 \oplus \cdots \oplus A_i^{n(i)}$, where $A_i^j (1 \le i \le n,\ 1 \le j \le n(i))$ are cyclically moving classes of symbols. We take $\Omega_i = \pi(X_{A_i})\ (1 \le i \le n)$ and $\Omega_i^j = \pi(X_{A_i^j})$ $(1 \le j \le n(i))$ (cf. [16] for more details).

The original proof of Proposition (III.1) by Smale involved stable manifold theory.

§1.2. Hyperbolic and Axiom A flows. The above approach for Axiom A diffeomorphisms can be used for Axiom A flows by adapting these constructions to the Poincaré map on certain transverse sections to the flow. Let M be a compact C^∞ manifold and $\varphi_t : M \to M$ a C^1 Axiom A flow.

DEFINITION. We call a point $x \in \Omega$ *wandering* if there exists an open neighbourhood U of x such that $\varphi_t U \cap U = \emptyset$ for all sufficiently large $t > 0$. The non-wandering set Ω is the complement of the union of the wandering points and is closed and φ-invariant.

The flow φ satisfies *Axiom A* if

(a) Ω is *hyperbolic*, i.e. there exists a continuous splitting $T_\Omega M = E^0 \oplus E^u \oplus E^s$ into a Whitney sum of $D\varphi$-invariant sub-bundles and there exist $C > 0, \lambda > 0$ such that:

$$\|D\varphi_t(v)\| \leq Ce^{-\lambda t}\|v\|, \text{ for } v \in E^s, t \geq 0;$$

$$\|D\varphi_{-t}(v)\| \leq Ce^{-\lambda t}\|v\|, \text{ for } v \in E^u, t \geq 0$$

and E^0 is one dimensional and tangent to the orbits of φ.

(b) The closed orbits are dense in Ω.

(III.5) PROPOSITION (Smale spectral decomposition). *We can decompose* $\Omega = \bigcup_{i=1}^{n} \Omega_i$, *where* Ω_i *are closed,* φ-invariant *disjoint sets and* $\varphi \mid \Omega_i$ *is transitive.*

The original proof used stable manifold theory (cf. [15]).

The above proposition allows us to restrict attention to $\varphi_t : \Omega_i \to \Omega_i$.

More generally, we can consider any differentiable flow on a closed invariant set $\Lambda \subset M$ such that:

(a) $\varphi_t : \Lambda \to \Lambda$ is hyperbolic;

(b) $\varphi_t : \Lambda \to \Lambda$ is transitive;

(c) the periodic orbits of $\varphi|_\Lambda$ are dense in Λ;

(d) there exists a neighbourhood $U \supset \Lambda$ with $\Lambda = \bigcap_{t=-\infty}^{\infty} \varphi_t(U)$.

We call $\varphi_t : \Lambda \to \Lambda$ a *hyperbolic flow*, and our principle examples are $\varphi_t : \Omega_i \to \Omega_i$. Henceforth we shall assume that Λ is neither a single closed orbit or a fixed point.

STABLE AND UNSTABLE MANIFOLDS AND LOCAL PRODUCT STRUCTURE. As for the case of diffeomorphisms the hyperbolic splitting of $T_\Lambda M$ under the flow gives rise to stable and unstable manifolds. An additional 1-dimensional sub-manifold is contributed by E^0.

For $\varepsilon > 0$ we define the (local) *stable manifold* for $x \in \Lambda$ by

$$W^s_\varepsilon(x) = \{y \in M : d(\varphi_t x, \varphi_t y) \le \varepsilon, \text{ for all } t \ge 0 \text{ and } d(\varphi_t x, \varphi_t y) \to 0 \text{ as } t \to +\infty\}$$

and the (local) *unstable manifold* for $x \in \Lambda$ by

$$W^u_\varepsilon(x) = \{y \in M : d(\varphi_{-t} x, \varphi_{-t} y) \le \varepsilon, \text{ for all } t \ge 0 \text{ and } d(\varphi_{-t} x, \varphi_{-t} y) \to 0 \text{ as } t \to +\infty\}.$$

For $\varepsilon > 0$ sufficiently small these form C^1 embedded discs with $T_x W^s_\varepsilon(x) = E^s_x$ and $T_x W^u_\varepsilon(x) = E^s_x$ (cf. [39] and [90]).

The *local product structure* for hyperbolic flows refers to the following property: for every $\delta > 0$ there exists $\eta > 0$ such that whenever $x,y \in \Lambda$ with $d(x,y) < \eta$ there exists unique $|t| \leq \delta$ such that $W^s_\varepsilon(\varphi_t x) \cap W^u_\varepsilon(y) \neq \emptyset$. Furthermore, the intersection is a single point of Λ which we denote by $\langle x,y \rangle$.

MARKOV SECTIONS. The basic idea is to construct transverse sections for the flow which have a special Markovian property. We require that the Poincaré map (induced by the flow) on these sections should transform them in a similar way to that of a hyperbolic diffeomorphism and a Markov partition.

We choose C^1 transverse sections $\{D_i\}_{i=1}^n \subseteq M$ and subsets $S_i \subset \operatorname{int} T_i \subset T_i \subset (\operatorname{int} D_i) \cap \Lambda$. We assume that $\bigcup_{i=1}^n \varphi_{[0,\alpha]} S_i = \Lambda$, for some real number $\alpha > 0$ and we choose $\eta = \eta(\delta)$ according to $\delta << \operatorname{dist}_{i \neq j}(D_i, D_j)$. If $(x,y) \in T_i \times T_i$ the image $\langle x,y \rangle$ need *not* lie in T_i. However we may project $\langle x,y \rangle$ to $[x,y] \in T_i$ along the orbits of the flow (providing α is sufficiently small).

DEFINITION. A set $R \subseteq T_i$ is a *rectangle* if whenever $x,y \in R$ then $[x,y] \in R$ and *proper* if $R = \overline{(\operatorname{int} R)}$.

We want to construct sections $\{R_j\} \subset \bigcup_{i=1}^n T_i$ which satisfy $\bigcup_j \varphi_{[0,\alpha]} R_j = \Lambda$, say, and $(\operatorname{int} R_i) \cap (\operatorname{int} R_j) = \emptyset$ for $i \neq j$.

Generically, a point $x \in \Lambda$ will generate a sequence in $\prod_{-\infty}^{\infty} \{1,\dots,k\}$ as it traverses sections under the flow. Here $H^n x \in \operatorname{int} R_{x_n}$, where $H : \bigcup_j R_j \to \bigcup_j R_j$ is the Poincaré map. In order that these sequences correspond to a subshift of finite type we impose an additional condition.

DEFINITION. The proper rectangles $\{R_i\}_{i=1}^{k}$ are *Markov sections* for $\varphi_t : \Lambda \to \Lambda$ if

(a) For $x \in \operatorname{int} R_i$ with $Hx \in \operatorname{int} R_j$ then $H(W^s(x,R_i)) \subset W^s(Hx,R_j)$, and

(b) For $x \in \operatorname{int} R_i$ with $H^{-1}x \in \operatorname{int} R_k$ then $H^{-1}(W^u(x,R_i)) \subset W^u(H^{-1}x,R_k)$,

where we write $W^s(x,R_j)$, $W^u(x,R_j)$ for the projections of $W^s_\varepsilon(x)$, $W^u_\varepsilon(x)$ onto $\bigcup_{j=1}^{k} R_j$.

The Poincaré map is discontinuous, so it is preferable to replace H as follows. We choose $T >> \alpha$ such that $Ce^{-\lambda(T-\alpha)} << 1$. We can choose open sets $\{U_j\}$ such that $\bigcup_i S_i \subset \bigcup_j U_j \subset \bigcup_i T_i$ and each U_j is sufficiently small that $\varphi_T U_j \subset \varphi_{[0,\alpha]} S_{i(j)}$ and $\varphi_{-T} U_j \subset \varphi_{[0,\alpha]} S_{k(j)}$.

This gives maps $H_j^+ : U_j \to S_{i(j)}$ and $H_j^- : U_j \to S_{k(j)}$ (by projecting along orbits of the flow) which are continuous on their domains and 'hyperbolic', for suitable $1 \le i(j), k(j) \le k$. The final refinement is to replace $\{U_j\}$ by a smaller cover $\{V_j\}$ (whose diameter is small compared with the Lebesgue number of $\{U_j\}$). Then for each ℓ choose $j = j(\ell)$ with $V_\ell \subset U_{j(\ell)}$. This induces $H^+ : V_\ell \to S_{i(j)}$, $H^- : V_\ell \to S_{k(j)}$.

Working with these maps we can repeat the constructions of Proposition (III.2) to find rectangles $\{R_i\}$ which are Markovian with respect to $H^+, H^- : \coprod_i R_i \to \coprod_i R_i$. (The rectangles can be made disjoint by flowing backwards or forwards incrementally under the flow.) If H^+, H^- correspond to at most n iterates of the Poincaré map $H : \coprod_i R_i \to \coprod_i R_i$ we may replace $\{R_i\}$ by $\{H^{-n}R_{i_{-n}} \cap \cdots \cap R_{i_0} \cap \cdots \cap H^n R_{i_n}\}$ (again made disjoint, if necessary, by flowing for an increment of time). These final sections are Markovian (with respect to H).

(III.6) PROPOSITION. *For a hyperbolic flow $\varphi_t : \Lambda \to \Lambda$ there exist (arbitrarily small) Markov sections for the flow.*

The only minor complication in the above proof is that in constructing $\{R_i\}$ from $\{V_\ell\}$ the new Markov sections may interfere with the Poincaré map on the old sections, i.e. encroach on the area between sections. (This would complicate the final step.) However, by a few extra technical assumptions this possibility can be eliminated (cf. [15] for full details).

We define a subshift of finite type $\sigma : X_A \to X_A$ by the $k \times k$ matrix

$$A(i,j) = \begin{cases} 1 & \text{if } H(\mathrm{int}\,R_i) \cap (\mathrm{int}\,R_j) \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

and define $\pi : X_A \to \bigcup_i T_i$ by $\pi(x) = \bigcap_{n=-\infty}^{\infty} \overline{H^{-n}(\mathrm{int}R_{x_n})}$. By analogy with the diffeomorphism case:

(III.7) PROPOSITION. π *is a well-defined map.*

Let $r(x) = \inf \{t > 0 : \varphi_t\pi(x) \in R_{x_1}\}$ *for* $x \in X_A$. *We can define a suspended flow* $\sigma_t^r : X_A^r \to X_A^r$ *and extend* $\pi : X_A^r \to \Lambda$ *by* $\pi(x,t) = \varphi_t\pi(x)$.

The effectiveness of σ_t^r in modelling φ is summarised in the following theorem.

(III.8) THEOREM (Bowen).

(i) π *is continuous and surjective*

(ii) π *is one-one on a set of full measure (for any ergodic measure of full support) and on a residual set*

(iii) π *is bounded-one*

(iv) $\pi\sigma_t^r = \varphi_t\pi$ *(for all* $t \in \mathbb{R}$).

The proof of this theorem parallels that of theorem (III.3), by working with $\sigma : X_A \to X_A$ and $H : \coprod_i T_i \to \coprod_i T_i$. (For full details, cf. [15].)

Since $\varphi_t : \Lambda \to \Lambda$ is transitive it is easy to see $\sigma : X \to X$ is transitive.

§2. Zeta-functions. To construct meromorphic extensions of zeta-functions for hyperbolic systems it is convenient to work at the level of symbolic dynamics. We can define zeta-functions for the symbolic systems which are explicitly related to the zeta-functions for the hyperbolic systems they model. By proving results on their domains at the symbolic system level we can infer results about their domain for the hyperbolic system.

As before, it is instructive to study the diffeomorphism case before considering the situation for flows.

§2.1. Zeta-functions for hyperbolic diffeomorphisms.

Let $f : \Lambda \to \Lambda$ be a hyperbolic diffeomorphism.

DEFINITION. The *zeta-function* for $f : \Lambda \to \Lambda$ is the complex function

$$\zeta(z) = \exp \sum_{n=1}^{\infty} z^n \frac{\mathrm{Card}\,\{x : f^n x = x\}}{n}, \quad z \in \mathbb{C}.$$

(This converges to a non-zero analytic function for $|z| < e^{-h}$, where h is the topological entropy of $f : \Lambda \to \Lambda$ as explained in Chapter 6.)

The *zeta-function* for $\sigma : X_A \to X_A$ is the complex function defined by

$$\zeta(z) = \exp \sum_{n=1}^{\infty} z^n \frac{\mathrm{Card}\,\{x : \sigma^n x = x\}}{n}, \quad z \in \mathbb{C}.$$

(The zeta-function is again well-defined for $|z| < e^{-h}$.)

The zeta-function for $\sigma : X_A \to X_A$ is clearly a simpler object since

$$\mathrm{Card}\,\{x : \sigma^n x = x\} = \mathrm{trace}\,(A^n) \text{ and so } \zeta(z) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \mathrm{trace}\,(A^n) = \frac{1}{\det(I-zA)}.$$

In general there is *not* a one-one correspondence between closed orbits for f and σ (and so we cannot expect to identify the zeta-functions for σ and f). The problem arises from periodic points for f lying on the boundaries of Markov partitions. However, Manning produced a combinatorial argument to account for these (see for example [33], [90]). Assume $f^n x = x \in T_{i_1} \cap \cdots \cap T_{i_m}$ then x is the image under π of distinct periodic points $x^1, \ldots, x^m$ with periods $N_1, \ldots, N_m$ with $n | N_i (i = 1, \ldots, m)$. Furthermore $\left\{x_n^i\right\}_{i=1}^m = \left\{x_0^i\right\}_{i=1}^m$.

(III.9) LEMMA. *For each* $j \in \mathbb{Z}$, $x_j^r \neq x_j^s$ *where* $1 \leq r < s \leq m$.

The proof is very similar to that of Theorem (III.3)(iii), except now $m \leq k$.

In particular, although $f^n x = x$, we can only deduce that the (ordered) family of rectangles $R_{x_n^j}$ $(j = 1,\ldots,m)$ is a permutation of $R_{x_0^j}$ $(j = 1,\ldots,m)$. Let γ denote the associated permutation and $\gamma = (\gamma_1,\ldots,\gamma_d)$ be the decomposition into cycles. (Clearly the number of x^j of period exactly n is precisely the number of 1-cycles.) Cycles of different lengths correspond to x^j whose periods are multiples of n. To compensate for this we need to introduce *more* subshifts. We let $\tau \subseteq 2^{\{T_1,\ldots,T_k\}}$ denote an (unordered) set of rectangles with non-empty intersection and $|\tau|$ the number of rectangles in τ. For pairwise disjoint $\tau_1,\ldots,\tau_n$, with $\tau_1 \cup \cdots \cup \tau_n$ containing rectangles with non-empty intersection (i.e. the rectangles in the union contain a common point), we denote $\hat{\tau} = (\tau_1,\ldots,\tau_n)$.

Given an n-tuple of positive integers $\underline{i} = (i_1,\ldots,i_n)$, where $n = n_{(\underline{i})}$, with $|\underline{i}| = i_1 + \cdots + i_n \leq k$ we want to define a subshift $\sigma^{(\underline{i})} : X_{A^{(\underline{i})}} \to X_{A^{(\underline{i})}}$ whose symbols are elements $\hat{\tau}$ with $|\tau_j| = i_j$, $1 \leq j \leq n$. The matrix $A^{(\underline{i})}$ is given by:

$A^{(\underline{i})}(\hat{\tau},\hat{\tau}')=$

$$\begin{cases} 1 & \text{if for each } 1 \le j \le n, \text{ there exists a permutation } \mu_j \text{ with } A\left(R_i^j, R^{j'}_{\mu_j(i)}\right) = 1 \\ & \qquad (1 \le i \le i_j), \text{ where } \tau_j = \left\{R_1^j,\ldots,R_{i_j}^j\right\}, \ \tau'_j = \left\{R_1^{j'},\ldots,R_{i_j}^{j'}\right\} \\ 0 & \text{otherwise.} \end{cases}$$

There is a unique map $\pi_{(\underline{i})} : X_{A(\underline{i})} \to \Lambda$ corresponding to $\pi : X_A \to \Lambda$.

(III.10) LEMMA (Manning).

For $f^n y = y \in \Lambda$, $\sum_{\underline{i}} (-1)^{n(\underline{i})+1} \, Card\left\{(\sigma^{(\underline{i})})^n x = x \in X_{A(\underline{i})} : \pi_{(\underline{i})} x = y\right\} = 1.$

This only involves pairing indices $\underline{i} = (i_1,\ldots,i_n)$ and $\underline{i}' = (i_1,\ldots,i_n, m - |\underline{i}|)$. This gives a cancellation (because of the difference in sign of $(-1)^{n(\underline{i})+1}$, $(-1)^{n(\underline{i}')+1}$) except for $\underline{i} = (m)$, which contributes the 1.

The next proposition follows directly from this lemma.

(III.11) PROPOSITION (Manning). $\zeta_f(z) = \left[\prod_{\substack{|\underline{i}| \text{ odd} \\ \underline{i} \neq (m)}} \zeta_{\sigma^{(\underline{i})}}(z) \Big/ \prod_{|\underline{i}| \text{ even}} \zeta_{\sigma^{(\underline{i})}}(z) \right] \zeta_\sigma(z)$. *In particular,* $\zeta_f(z)$ *has a meromorphic extension to* $\mathbb{C}$ *as a rational function.*

There is a slightly simplified version of the proof above (due to Bowen) given in ([18]).

§2.2. Zeta functions for hyperbolic flows.

Let $\varphi_t : \Lambda \to \Lambda$ be a hyperbolic flow restricted to its non-wandering set. Let $\sigma^r_t : X^r_A \to X^r_A$ be a suspended flow modelling φ.

DEFINITION. The zeta function $\zeta_{\sigma^r}(s)$ for $\sigma^r_t : X^r_A \to X^r_A$ is the complex function $\zeta_{\sigma^r}(s) = \prod_\tau (1-e^{-s\lambda(\tau)})^{-1}$, $s \in \mathbb{C}$, where the Euler product is over all closed orbits τ for σ^r_t. (This converges to a non-zero analytic function of s for $\mathcal{R}(s) > h$, where h is the topological entropy of σ^r.)

One may also write this as

$$\zeta_{\sigma^r}(s) = \exp - \sum_\tau \log(1-e^{-s\lambda(\tau)}) = \exp \sum_\tau \sum_{n=1}^{\infty} (e^{-s\lambda(\tau)})^n/n .$$

The zeta function for $\varphi_t : \Lambda \to \Lambda$ is defined similarly. Each σ^r-periodic orbit τ corresponds to a periodic orbit $\{x, \sigma x, \ldots, \sigma^{n-1}x\}$ of least period n, say, with $\lambda(\tau) = r^n(x) = r(x) + r(\sigma x) + \cdots + r(\sigma^{n-1}x)$.

$$\text{Therefore, } \zeta_{\sigma^r}(s) = \exp \sum_{k=1}^{\infty} \sum_{\substack{\sigma^k x = x \\ k = \text{least period}}} \frac{1}{k} \left(\sum_{n=1}^{\infty} \frac{e^{sr^k(x)n}}{n} \right)$$

$$= \exp \sum_{m=1}^{\infty} \frac{1}{m} \sum_{\sigma^m x = x} e^{sr^m(x)n}$$

(where we take $m = kn$).

As in the case of diffeomorphisms there is not a one-one correspondence between closed orbits for φ and σ^r, and so ζ_{σ^r} and ζ_φ cannot be immediately identified. The difficulty arises with φ periodic orbits passing through the boundaries of Markov sections. Bowen showed how Manning's combinatorial argument for the discrete case could be modified for flows. (There are extra complications for flows over and above those for diffeomorphisms. We want to use the Poincaré map on sections to apply Manning's lemma. The disjointness of the sections suggests the need for a slightly more involved construction.)

For each rectangle $R_i (i = 1,\ldots,k)$ let $P_i = \{\varphi_t \pi(x) : t \in [0, r(x)),\ x_0 = i\}$ be the "parallelogram" swept out by R_i. Given an index $\underline{i} = (i_1,\ldots,i_n)$ (an n-tuple of positive integers with $|\underline{i}| = i_1 + \cdots + i_n \leq k$) we define a set of symbols $\hat{\tau} = (R_\alpha ; \tau_1,\ldots,\tau_n)$ with $\tau_1,\ldots,\tau_n \subset 2^{\{R_1,\ldots,R_k\}}$, disjoint, but whose sets have a common intersection, and Card $\tau_j = i_j (j = 1,\ldots,n)$, with $R_\alpha \subset \bigcup_{j=1}^{n} \tau_j$. Furthermore, we require that R_α, be the 'leading rectangle' in the natural sense. We define a transition matrix indexed by $\hat{\tau}$ by

$$A^{(\underline{i})}(\hat{\tau},\hat{\tau}') = \begin{cases} 1 & \text{if } \bigcup_{j=1}^{n} \tau_j \,,\, \bigcup_{j=1}^{n} \tau'_j \text{ differ by } R_\beta , R'_\gamma \text{ with } A(R_\beta , R'_\gamma) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

There is a corresponding subshift $\sigma^{\underline{i}} : X_{A(\underline{i})} \to X_{A(\underline{i})}$ and induced maps $r_{\underline{i}} : X_{A(\underline{i})} \to \mathbb{R}^+$, $\pi_{\underline{i}} : X_{A(\underline{i})} \to \Lambda$.

(III.12) LEMMA (Bowen, after Manning). *For a closed* φ-orbit τ *of least period* ℓ *we have*

$$1 = \sum_{\underline{i}} (-1)^{n(\underline{i})+1} \, \mathit{Card}\, \{\tau_{\underline{i}} : \pi_{\underline{i}}(\tau_{\underline{i}}) = \tau \,,\, \lambda(\tau_{\underline{i}}) = \ell\}$$

(where $\tau_{\underline{i}}$ *denotes a closed orbit for* $\sigma_{\underline{i}}^{r_{\underline{i}}}$*).*

The proof is essentially the same as that of Manning's lemma (except that there are slightly more details to pay attention to).

The following result follows directly from the above lemma.

(III.13) PROPOSITION (Bowen, after Manning).

$$\zeta_\varphi(s) = \Big(\prod_{\substack{|\underline{i}|\ \mathrm{odd} \\ \underline{i} \neq (m)}} \zeta_{\sigma^{(\underline{i})}}(s) \Big/ \prod_{|\underline{i}|\ \mathrm{even}} \zeta_{\sigma^{(\underline{i})}}(s) \Big) \zeta_\sigma(s)\,,$$

where $\zeta_\sigma(s)$ *corresponds to* $\sigma^{(\underline{i})}$ *with* $\underline{i} = (m)$.

(We refer the reader to ([15]) for full details of proofs.)

Notes

§1.1. Anosov diffeomorphisms were originally introduced by D.V. Anosov, under the name of C-diffeomorphisms, in 1962 [5]. Smale proposed the generalisation to Axiom A diffeomorphisms, see for example his 1967 survey paper [95].

The stable manifolds and local product structure were examined in a series of papers by Hirsch, Pugh, Shub, *et al.,* culminating in their 1975 book [39].

The construction of Markov partitions for Anosov diffeomorphisms was done by Sinai in 1968 [92]. This followed the highly illustrative but special case of two-dimensional toral automorphisms studied by Adler and Weiss. The generalisation to Axiom A diffeomorphisms is due to Bowen [10]. There is an alternative approach using the 'shadowing property' in Bowen's 1975 book [16]. Ruelle proposed a further generalisation to Smale spaces in his 1978 *Thermodynamic Formalism* book.

§1.2. Anosov flows were introduced by Anosov (as C-flows) and were studied extensively by him in his thesis, published in English in 1967 [5]. They were intended to be generalisations of geodesic flows on compact surfaces with strictly negative sectional curvatures. Smale proposed the more general Axiom A flows. See, for example, his survey paper [95].

The stable manifold theory and local product structure are dealt with in the work of Hirsch, Pugh, Shub [39].

For 3-dimensional manifolds the symbolic dynamics for Anosov flows were constructed by M. Ratner in 1969 [73]. The generalisation to any dimension by the same author followed in 1973 [75]. Bowen's construction of symbolic dynamics for Axiom A flows, extending his own work on Axiom A diffeomorphisms, appeared in the same year [15].

§2. The question of rationality of $\zeta(z)$ for Axiom A diffeomorphisms was originally posed by Smale in his article [95]. The problem was completely solved by A. Manning in his 1972 Ph.D. thesis. Earlier partial results, include those by Guckenheimer and Williams. For topological approaches, based on some form of Lefschetz fixed point theorem, see for example Franks [32] (cf. also Fried's paper [33]). The extension of Manning's proof to flows is due to Bowen [15].

APPENDIX IV

GEODESIC FLOWS

Probably the single most important example of an Axiom A flow is the geodesic flow on the unit tangent bundle of a compact manifold with strictly negative sectional curvatures. We now want to give some indication as to why these flows satisfy Axiom A (where the entire unit tangent bundle is the non-wandering set).

Assume M is an n-dimensional C^∞ compact Riemannian manifold whose Riemannian metric $\langle , \rangle$ has strictly negative sectional curvatures. The *geodesic flow* $\varphi_t : T_1M \to T_1M$ (on $T_1M = \{(x,v) \in TM : \langle v,v \rangle_x = 1\}$) is defined as follows:

Given $(x,v) \in T_1M$ let $\gamma : \mathbb{R} \to M$ be the unique unit speed geodesic through $x \in M$ in the direction v at time $t = 0$ (i.e. $\gamma(0) = x$, $\dot{\gamma}(0) = v$) then set $\varphi_t(x,v) = (\gamma(t),\dot{\gamma}(t))$. (Thus φ_t moves the tangent vector from $\gamma(0)$ to $\gamma(t)$ along the geodesic determined by v.)

To establish hyperbolicity we need a better understanding of $D\varphi_t : T(T_1M) \to T(T_1M)$. For convenience we shall consider TM rather than T_1M.

The map $\pi : TM \to M$ given by $\pi(x,v) = x$ has a derivative $D\pi_{(x,v)}$: $T_{(x,v)}(TM) \to T_xM$. We can define a second (linear) map $K : T_{(x,v)}(TM) \to T_xM$

by first choosing for $\xi \in T_{(x,v)}(TM)$ a curve $Z : (-\varepsilon,\varepsilon) \to TM$ tangent to ξ at time $t = 0$. For the composite curve $\alpha = \pi \circ Z : (-\varepsilon,\varepsilon) \to M$, say, we can set $K\xi = \nabla_\alpha Z$ (i.e. the covariant derivative of Z along α at time $t = 0$).

For each $(x,v) \in TM$ we can decompose $T_{(x,v)}(TM) = (\mathrm{Ker}\, D\pi) \oplus (\mathrm{Ker}\, K)$ (cf. [5] for a detailed account).

Using the Riemannian metric $\langle,\rangle$ for M we can define a metric for TM by $\langle\xi,\eta\rangle_{(x,v)} = \langle D\pi\xi, D\pi\eta\rangle_x + \langle K\xi, K\eta\rangle_x$, for $\xi,\eta \in T_{(x,v)}(TM)$.

Returning to the flow, we have for every point $(x,v) \in TM$ an associated geodesic γ with $\gamma(0) = x$, $\dot\gamma(0) = v$. Given $\xi \in T_{(x,v)}(TM)$ we associate with it the Jacobi field Y_ξ along γ such that $Y_\xi(x) = D\pi\xi$, $\nabla_\gamma Y_\xi(x) = K\xi$. The map $\xi \mapsto Y_\xi$ is a linear isomorphism from $T_{(x,v)}(TM)$ to Jacobi fields on γ (where γ is determined by (x,v)).

The derivative of the flow $D\varphi_t : T(TM) \to T(TM)$ is described by $D\pi[D\varphi_t(\xi)] = Y_\xi(\gamma(t))$ and $K[D\varphi_t(\xi)] = (\nabla_\gamma Y_\xi)(\gamma(t))$.

To check the hyperbolicity condition it is convenient to introduce an adapted frame field for the geodesic γ i.e. a system of parallel orthogonal vector fields $e_1(t),\dots,e_n(t)$ along γ with γ tangent to $e_n(t)$ at $t = 0$. A C^∞ vector field Y on γ is identified with an $(n-1)$-tuple $(y_1(t),\dots,y_{n-1}(t))$ of C^∞ functions $y_i : \mathbb{R} \to \mathbb{R}$ $(i=1,\dots,n-1)$ by $Y(\gamma(t)) = \sum_{i=1}^{n-1} y_i(t)e_i(t)$ and $\nabla_\gamma Y(\gamma(t)) = \sum_{i=1}^{n-1} y'_i(t)e_i(t)$.

The sectional curvatures of M (with respect to $\langle , \rangle$) have an important influence on Jacobi fields Y, and hence on $D\varphi_t$. Let R denote the curvature tensor for M. (Recall $R(X,Y)Z = [\nabla_X Y, \nabla_Y X] - \nabla_{[X,Y]}Z$.) We can define an $(n-1) \times (n-1)$ matrix by $R_{ij}(t) = \langle R(e_n(t), e_i(t))\, e_n(t), e_j(t)\rangle$, for each $t \in \mathbb{R}$. A perpendicular Jacobi vector field on γ is defined by $t \mapsto Y(\gamma(t))x$, where $x \in \mathbb{R}^{n-1}$ and $Y(\gamma(t))$ is a solution of the $(n-1) \times (n-1)$ matrix differential equation

$$Y''(\gamma(t)) + R(t)Y(\gamma(t)) = 0.$$

(For a surface, this reduces to a single differential equation, where $R(t)$ is the curvature of the surface at $\gamma(t)$.)

Since the solutions to this equation are uniquely determined by the initial conditions, there are $2(n-1)$ independent perpendicular Jacobi vector fields.

The entries R_{ij} for the matrix R are sectional cuvatures for the manifold M, which by hypothesis are all strictly negative. This is the main point since the solutions either decay exponentially fast (in $(n-1)$ dimensions) or blow up exponentially fast (in $(n-1)$ dimensions) cf. [5]. In our previous discussion this corresponds to the hyperbolicity of $\varphi_t : T_1M \to T_1M$.

CONSTANT CURVATURE AND FUCHSIAN GROUPS. There are alternative, somewhat more canonical, ways of constructing suspended flows in the case of geodesic flows for compact surfaces of constant negative curvature. The origins of this approach lie in the work of Morse and Nielsen [59]. The refined version we shall describe is due to Series some of the initial steps having evolved from joint work with Bowen. Cf. [19], [87]. Closely related work is due to Adler and Flatto [4].

Assume that S is a compact surface of constant curvature $\kappa = -1$. Let $\varphi_t : N \to N$ be the associated geodesic flow on the unit tangent bundle $N = T_1S$. The universal cover $\tilde{S}$ for S, with the metric lifted from S, can be identified with the interior of the unit disc $D = \{z \in \mathbb{C} : |z| < 1\}$ with the Poincaré metric $ds^2 = \frac{1}{4}(dx^2+dy^2)/(1-(x^2+y^2))^2$, where $z = x+iy$. The deck transformations for the projection $\pi : D \to S$ form a discrete group Γ of orientation preserving isometries of (D,ds). Such a group is called a Fuchsian group. For the Poincaré metric every isometry takes the special form of a linear fractional transformation of the type $g : z \mapsto (az+b)/(\bar{b}z+\bar{a})$ where $a,b \in \mathbb{C}$, $|a|^2-|b|^2 = 1$.

The geodesics in (D,ds) take an especially simple form. As a point-set they are circular arcs in D which meet the unit circle $K = \{z : |z| = 1\}$ perpendicularly. In particular, a pair of distinct points $(x,y) \in K \times K - \text{diag.}(K \times K)$ determine a unique (directed) geodesic γ in D by specifying its asymptotic points, i.e. $\gamma(+\infty) = y$, $\gamma(-\infty) = x$.

Clearly we can identify $S = D/\Gamma$. However, there is a canonical 'copy' of S in D.

For any $g \in \Gamma$ we call $C(g) = \{z \in D : |g'(z)| = 1\}$ the isometric circle of g. (This set is a geodesic arc in D.) Since $(g^{-1}g)'(z) = 1 = (g^{-1})'(gz).g'(z)$ we observe that $gC(g) = C(g^{-1})$. There exists a (non-unique) special choice of (finitely many) generators $\Gamma_0 \subset \Gamma$ such that, in particular, the compact region of R exterior to all of the arcs $\{C(g) : g \in \Gamma_0\}$ represents a copy of S in D (with piecewise geodesic boundaries). R is called a *fundamental region* of Γ.

Let γ denote a directed geodesic on S then γ will have many lifts to D. Assume, for the sake of argument, we choose a lift $\tilde{\gamma}$ on D with $\tilde{\gamma} \cap \bar{R} \neq \emptyset$. Let $\tilde{\gamma}$ have base points $(x,y) \in K \times K$ (i.e. $\tilde{\gamma}$ is forward asymptotic to y, say, and backward asymptotic to x).

The geodesic flow $\varphi_t : T_1S \to T_1S$ 'lifts' to a geodesic flow $\tilde{\varphi}_t : T_1D \to T_1D$ on the unit tangent bundle T_1D of D. Consider the action of $\tilde{\varphi}_t$ on the lifted geodesic $\tilde{\gamma}$. If v is a tangent vector for $R \cap \tilde{\gamma}$ then the geodesic flow $\tilde{\varphi}_t$ will transport v to the boundary of R and then into a new region gR, $g \in \Gamma_0$, say.

Observe that the action of g^{-1} moves this region back to R and that $\tilde{\gamma}$ is replaced by $g^{-1}\tilde{\gamma}$ with new base points $(g^{-1}x, g^{-1}y)$. The 'Markov partition' and 'symbolic dynamics' we want to describe are association with this action on $K \times K$.

Series shows how to construct a (finite) partition of K into intervals (or arcs) $\{I(g) : g \in \Gamma_0\}$ (in particular, $I(g)$ lies in the part of K interior to $C(g)$) and defines $f^+ : K \to K$ by $f^+ \mid I(g) = g$. (There is a minor ambiguity at the finite set of endpoints.) Without loss of generality assume that R was chosen with boundary pieces which meet perpendicularly (cf. [89]), then relative to a suitable sub-partition $\{I_i\}_{i=1}^k$ of $\{I(g) : g \in \Gamma_0\}$ the endomorphism $f^+ : K \to K$ becomes Markov, and we can write $y = \bigcap_{n=0}^{\infty} (f^+)^{-n} I_{y_n}$, say. Similarly, we can define a second endomorphism $f^- : K \to K$ by $f^- \mid J(g) = g$ (for a partition $\{J(g) : g \in \Gamma_0\}$) which is Markov relative to some subpartition $\{J_i\}_{i=1}^k$ of $\{J(g) : g \in \Gamma_0\}$, and we can write $x = \bigcap_{n=0}^{\infty} (f^-)^{-n} J_{x_n}$.

The sequences $\underline{z} = (\ldots, x_2, x_1, x_0, y_0, y_1, \ldots)$ correspond to a (two-sided) shift X_A. For $\underline{z} \in X_A$ we can associate the directed geodesic $\tilde{\gamma}$ with base points (x,y). For "most" geodesics $\tilde{\gamma} \cap \bar{R} \neq \emptyset$ and we define $r(\underline{z}) = \text{length}\,(\tilde{\gamma} \cap \bar{R}) \in \mathbb{R}$ and let $\pi(\underline{z}) \in T_1 D \mid \partial R$ be the tangent vector to $\tilde{\gamma}$ as it enters R. (For a detailed account of all cases, including the case $\tilde{\gamma} \cap \bar{R} = \emptyset$, cf. [89]). We extend to $\pi : X_A^r \to N$ by $\pi(\underline{x},t) = \varphi_t \pi(\underline{x})$ (with obvious identifications). Finally, we have the following analogue of Lemma 9.1:

(IV.1) THEOREM (Series)

(i) π *is continuous and surjective.*

(ii) π *is one-one on a set of full measure (for every ergodic measure of full support)*

(iii) π *is bounded-one (in fact, at most 4-to-one)*

(iv $\pi\sigma_t^r = \varphi_t\pi$.

Notes

The proof that geodesic flows for compact manifolds with negative sectional curvatures are Anosov was proved by D.V. Anosov in his thesis [5].

Coding geodesics by generators, for surfaces of constant negative curvature, has historical roots in the work of Nielsen, Hadamard, Koebe and Morse. The foundations for the ergodic theory of geodesics flows were laid by Hedlund [37] (see also [41]).

The systematic account we describe is basically due to C. Series. This started with a collaboration with Bowen, which was completed after his death in 1978 [19]. However, a version of the symbolic dynamics as we describe it did not appear until her later 1981 paper [87]. In another paper Series gives an alternative, and perhaps more appealing, way of constructing the intervals on K [89].

APPENDIX V

PERTURBATION THEORY FOR LINEAR OPERATORS

Here we present a brief account of the analytic perturbation theory referred to in earlier chapters. Complete details may be found in Kato's book [44] or (sufficient for our needs) in Bhatia and Parthasarathy's lecture notes [8].

Let B be a complex Banach space. A map $f : \mathbb{C} \to B$ is said to be *analytic* if $\ell \circ f : \mathbb{C} \to \mathbb{C}$ is analytic in the usual sense for any bounded linear functional $\ell : B \to \mathbb{C}$. If B_1, B_2 are complex Banach spaces then $g : B_1 \to B_2$ is said to be *analytic* if $g \circ f : \mathbb{C} \to B_2$ is analytic for any analytic map $f : \mathbb{C} \to B_1$. These notions may be localised and in particular one may define *real analyticity* for maps of open subsets of real Banach spaces into real Banach spaces.

Let $L : B \to B$ be a bounded linear operator on a complex Banach space (if B were a real Banach space we could take its complexification and extend L accordingly) and define

$$Sp(L) = \{\lambda \in \mathbb{C} : (\lambda I - L) \text{ is not invertible}\}$$

to be the *spectrum* of L.

If the closed bounded set $Sp(L) = \Sigma_1 \cup \Sigma_2$ is decomposed into disjoint non-empty sets Σ_1, Σ_2 and if Γ is a closed simple curve in $\mathbb{C}$ disjoint from $sp(L)$ which has Σ_1 in its interior and Σ_2 in its exterior then the bounded linear operator $\pi : B \to B$ given by

$$\pi = -\frac{1}{2\pi i}\int_\Gamma (z-L)^{-1}dz$$

is a *projection* i.e. $\|\pi\| = 1$ and $\pi^2 = \pi$. Moreover we can write $B = B_1 \oplus B_2$ where $B_1 = \pi(B)$ $B_2 = (I-\pi)B$ are closed L invariant subspaces and $sp(L|B_1) = \Sigma_1$, $sp(L|B_2) = \Sigma_2$.

Assume now that $\Sigma_1 = \{\lambda\}$ consists of a single simple eigenvalue isolated (by Γ) from the rest of the spectrum of L then for any bounded linear operator $L' : B \to B$ sufficiently close to L the spectrum of L' may be written $sp(L') = \Sigma'_1 \cup \Sigma'_2$ where $\Sigma'_1 = \{\lambda'\}$ consists of a single simple eigenvalue isolated from Σ'_2 by Γ. The projection π associated with $\Sigma_1 = \{\lambda\}$ is called the *eigenprojection* of λ and the map $L' \to \pi'$ which associates the eigenprojection to the operator is analytic in a neighbourhood of L.

As a consequence the map $L' \to \lambda'$ which associates the eigenvalue λ' to the operator L' is analytic in a neighbourhood of L. A further consequence is that Σ'_2 remains within a preassigned neighbourhood of Σ_2 if L' is sufficiently close to L.

Finally we remark that if λ is an isolated eigenvalue of L of finite multiplicity n, then for sufficiently close operators L' the spectrum of L' within

Γ will consist of eigenvalues $\lambda'_1,\ldots,\lambda'_n$ and $L' \to \pi'$ will again be analytic. However one cannot assert that individual eigenvalues are analytically dependent on L' but only that $\text{Trace } L' = \lambda'_1+\cdots+\lambda'_n$ and $\det L' = \lambda'_1\cdots\lambda'_n$ are analytic.

REFERENCES

[1] L.M. Abramov, The entropy of a derived automorphism, *Dokl. Akad. Nauk. SSSR,* 128 (1959), 647-650.

[2] L.M. Abramov, On the entropy of a flow, *Dokl. Akad. Nauk. SSSR,* 128 (1959), 873-875.

[3] T. Adachi and T. Sunada, Homology of closed geodesics in a negatively curved manifold, *J. Diff. Geom.,* 26 (1987), 81-99.

[4] R. Adler and L. Flatto, Geodesic flows, interval maps, and symbolic dynamics, Research report, I.B.M., Yorktown Heights.

[5] D. Anosov, Geodesic flows on closed Riemannian manifolds with negative curvature, *Proc. Steklov Inst. Math.,* 90 (1967), 1-235.

[6] D. Anosov and Y.G. Sinai, Some smooth ergodic systems, *Russian Math. Surv.,* 22 (1967), 103-167.

[7] E. Artin and B. Mazur, On periodic points, *Annals of Math.,* 81 (1965), 82-99.

[8] R. Bhatia and K.R. Parthasarathy, *Lectures on functional analysis, I,* MacMillan, Delhi, 1977.

[9] R. Bowen and O. Lanford, Zeta functions of restrictions of the shift transformation, *Proc. Symp. Pure Math.,* 14 (1970), 43-50.

[10] R. Bowen, Markov partitions for Axiom A diffeomorphisms, *Amer. J. Math.,* 92 (1970), 725-747.

[11] R. Bowen, One-dimensional hyperbolic sets for flows, *J. Diff. Equations,* 12 (1972), 173-179.

[12] R. Bowen, The equidistribution of closed geodesics, *Amer. J. Math.,* 94 (1972), 413-423.

[13] R. Bowen and P. Walters, Expansive one-parameter flows, *J. Diff. Eq.,* 12 (1972), 180-193.

[14] R. Bowen, Periodic orbits for hyperbolic flows, *Amer. J. Math.,* 94 (1972),

1-30.

[15] R. Bowen, Symbolic dynamics for hyperbolic flows, *Amer. J. Math.*, 95 (1973), 429-460.

[16] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, SLN 470, Springer, Berlin, 1975.

[17] R. Bowen and D. Ruelle, The ergodic theory of Axiom A flows, *Invent. Math.*, 29 (1975), 181-202.

[18] R. Bowen, *On Axiom A diffeomorphisms*, Reg. Conf. Series 35, Amer. Math. Soc., 1977.

[19] R. Bowen and C. Series, Markov maps associated to Fuchsian groups, *Pub. Math. (IHES)*, 50 (1979), 153-170.

[20] M. Brin, Ergodic theory of frame flows. In *Ergodic theory and dynamical systems*, II (ed. A. Katok), Birkhauser: Basel, 1982.

[21] F. Browder, On the spectral theory of elliptic differential operators, I, *Math. Ann.*, 142 (1961), 22-130.

[22] J.W.S. Cassels and A. Fröhlich, *Algebraic number theory*, Academic, London, 1967.

[23] Z. Coelho and W. Parry, Central limit asymptotics for subshifts of finite type, *Israel J. Math.*, 69, no. 2 (1990), 235-249.

[24] C. Corduneau, *Almost periodic functions*, Interscience, New York, 1968.

[25] I.P. Cornfield, S.V. Fomin and Y.G. Sinai, *Ergodic theory*, Springer, Berlin, 1983.

[26] H. Delange, Généralisation du théorème de Ikehara, *Ann. Sci. E.N.S.*, 71 (1954), 213-242.

[27] M. Denker and W. Phillip, Approximation by Brownian motion for Gibbs measures and flows under a function, *Ergod. Th. and Dynam. Sys.*, 4 (1984), 541-552.

[28] M. Denker, C. Grillenburger and K. Sigmund, *Ergodic theory on compact spaces*, SLN 527, Springer, Berlin, 1976.

[29] P. Eberlein, Geodesic flows on negatively curved manifolds, I, *Annals of Math.*, 95 (1972), 492-510.

[30] W. & F. Ellison, *Prime numbers,* Herman, Paris, 1975.

[31] W. Feller, *An introduction to probability theory and its applications,* vol. 2, New York, J. Wiley & Sons, 1971.

[32] J. Franks, *Homology and dynamical systems,* Reg. Conf. Series, 49, Amer. Math. Soc., 1982.

[33] D. Fried, Rationality for isolated expansive sets, *Adv. Math.*, 65 (1987), 35-38.

[34] G. Gallavotti, Funzioni zeta ed insiemi basilas, *Accad. Lincei, Rend. Sc. fismat. e nat.*, 61 (1976), 309-317.

[35] F. Gantmacher, *The theory of matrices,* vol. II, Chelsea, New York, 1974.

[35*] Y. Guivarc'h, Propriétés ergodiques, en mesure infinie, de certains systèmes dynamiques fibrés, *Ergod. Th. and Dynam. Sys.*, 9 (1989), 433-453.

[35**] Y. Guivarc'h and J. Hardy, Théorèmes limites pour une classe de chaînes de Markov et applications aux difféomorphisms d'Anosov, *Annales de l'Inst. H. Poincaré,* 24 (1) (1988), 73-98.

[36] N.T.A. Haydn, Meromorphic extension of the zeta function for Axiom A flows, *Ergod. Th. and Dynam. Sys.*, 10 (1990), 347-360.

[37] G. Hedlund, The dynamics of geodesic flows, *Bull. Amer. Math. Soc.*, 45 (1939), 241-246.

[38] D. Hejhal, *The Selberg trace formula for* PSL(2,$\mathbb{R}$), SLN 548, Springer, Berlin, 1976.

[39] M. Hirsch, C. Pugh and M. Shub, *Invariant manifolds,* SLN 583, Springer, Berlin, 1977.

[40] F. Hofbauer and G. Keller, Zeta-functions and transfer operators for piecewise linear transformaltions, *J. Reine Angew. Math.*, 352 (1984), 100-113.

[41] E. Hopf, *Ergodentheorie,* Springer, Berlin, 1937.

[42] H. Huber, Zur analytischen Theorie hyperbolischer Raum formen und Bewegungsgruppen, II, *Math. Ann.*, 142 (1961), 385-398.

[43] C. Ionescu-Tulcea and G. Marinescu, Théorie ergodique pour les classes d'opérations non completement continues, *Annals Math.*, 52 (1950), 140-147.

[44] T. Kato, *Perturbation theory of linear operators,* Springer, Berlin, 1966.

[45] A. Katsuda and T. Sunada, Dynamical L-functions and homology of closed orbits, *Bull. Amer. Math. Soc.*, 20, no. 1 (1989), 73-77.

[46] A. Katsuda and T. Sunada, Closed orbits in homology classes (Preprint).

[47] A. Katsuda and T. Sunada, Homology and closed geodesics in a compact Riemann surface, *Amer. J. Math.*, 109 (1987), 145-156.

[48] Y. Katznelson, *An introduction to harmonic analysis,* J. Wiley & Sons, New York, 1968.

[49] M. Keane, Strongly mixing g-measures, *Invent. Math.*, 16 (1972), 309-324.

[50] G. Keller, Un théorème de la limite centrale pour une classes de transformations monotone par morceaux, *CRAS,* A291 (1980), 155-158.

[51] G. Keller, On the rate of convergence to equilibrium in one-dimensional systems, *Comm. Math. Phys.*, 96 (1984), 181-193.

[52] S. Lalley, Distribution of periodic orbits of symbolic and Axiom A flows, *Adv. Appl. Math.*, 8 (1987), 154-193.

[53] O. Lanford and D. Ruelle, Observables at infinity and states with short range correlations in statistical mechanics, *Commun. Math. Phys.*, 13 (1969), 194-215.

[54] S. Lang, *Algebraic number theory,* Addison-Wesley, 1970.

[55] F. Ledrappier, Principe variationnel et systèmes dynamiques symboliques, *Z. Wahr. Verw. Geb.*, 30 (1974), 185-202.

[56] A. Livsic, Cohomology properties of dynamical systems, *Math. USSR-Izv.*, 6 (1972), 1278-1301.

[57] G. Margulis, On some applications of ergodic theory to the study of manifolds of negative curvature, *Func. Anal. App.*, 3 (1969), 89-90.

[58] M. Misiurewicz, A short proof of the variational principle for a Z_+^N action on a compact space, *Asterisque,* 40 (1976), 147-187.

[59] M. Morse, Symbolic dynamics, Mimeographed notes, Princeton, 1966.

[60] R. Nussbaum, The radius of the essential spectrum, *Duke Math. J.,* 37 (1970), 473-478.

[61] W. Parry, Intrinsic Markov chains, *Trans. Amer. Math. Soc.,* 112 (1964), 55-65.

[62] W. Parry, *Topics in ergodic theory,* C.U.P., Cambridge, 1981.

[63] W. Parry and S. Tuncel, *Classification problems in ergodic theory,* C.U.P., Cambridge, 1982.

[64] W. Parry and S. Tuncel, On the stochastic and topological structure of Markov chains, *Bull. London Math. Soc.,* 14 (1982), 16-27.

[65] W. Parry, An analogue of the prime number theorem for closed orbits of shifts of finite type and their suspensions, *Israel J. Math.,* 45 (1983), 41-52.

[66] W. Parry and M. Pollicott, An analogue of the prime number theorem and closed orbits of Axiom A flows, *Annals of Math.,* 118 (1983), 573-591.

[67] W. Parry, Bowen's equidistribution theory and the Dirichlet density theorem, *Ergod. Th. and Dynam. Sys.,* 4 (1984), 117-134.

[68] W. Parry and M. Pollicott, The Chebotarev theorem for Galois coverings of Axiom A flows, *Ergod. Th. and Dynam. Sys.,* 6 (1986), 133-148.

[69] W. Parry, Sychronisation of canonical measures for hyperbolic attractors, *Commun. Math. Phys.,* 106 (1986), 267-275.

[70] S. Patterson, The limit set of a Fuchsian group, *Acta Math.,* 136 (1976),

241-273.

[71] M. Pollicott, A complex Ruelle operator theorem and two counter examples, *Ergod. Th. and Dynam. Sys.*, 4 (1984), 135-146.

[72] M. Pollicott, Meromorphic extensions for generalised zeta functions, *Invent. Math.*, 85 (1986).

[73] M. Ratner, Markov partitions for Anosov flows on 3-dimensional manifolds, *Mat. Zam.*, 6 (1969), 693-704.

[74] M. Ratner, The central limit theorem for y-flows on three dimensional manifolds, *Dokl. Akad. Nauk. SSSR*, 186 (1969), 519-521.

[75] M. Ratner, The central limit theorem for geodesic flows on n-dimensional manifolds of negative curvature, *Israel J. Math.*, 16 (1973), 181-197.

[76] S.M. Rees, Checking ergodicity of some geodesic flows with infinite Gibbs measure, *Ergod. Th. and Dynam. Sys.*, 1 (1981), 107-133.

[76*] J. Rousseau-Egelé, Un théorème de la limite locale pour une classe de transformations dilatantes, *Ann. of Probability*, 11, no. 3 (1983), 772-788.

[77] D. Ruelle, Statistical mechanics of a one-dimensional lattice gas, *Commun. Math. Phys.*, 9 (1968), 267-278.

[78] D. Ruelle, Statistical mechanics on a compact set with $\mathbb{Z}^\nu$ action satisfying expansiveness and specification, *Trans. Amer. Math. Soc.*, 185 (1973), 237-253.

[79] D. Ruelle, Zeta functions for expanding maps and Anosov flows, *Invent. Math.*, 34 (1978), 231-242.

[80] D. Ruelle, Generalised zeta functions for Axiom A basic sets, *Bull. Amer. Math. Soc.*, 82 (1976), 153-156.

[81] D. Ruelle, A measure associated with Axiom A attractors, *Amer. J. Math.*, 98 (1976), 619-654.

[82] D. Ruelle, *Thermodynamic formalism*, Addison-Wesley, 1978.

[83] D. Ruelle, Flows which do not exponentially mix, *C.R.A.S.*, 296 (1983), 191-194.

[84] P. Sarnak, Prime geodesic theorems, Ph.D. thesis, Stanford, 1980.

[85] A. Selberg, Harmonic analysis and discontinuous subgroups in weakly symmetric Riemannian spaces with applications to Dirichlet series, *J. Indian Math. Soc.*, 20 (1956), 47-87.

[86] E. Seneta, *Non-negative matrices,* George Allen and Unwin, London, 1973.

[87] C. Series, Symbolic dynamics for geodesic flows, *Acta Math.*, 146 (1981), 103-128.

[88] C. Series, The infinite word problem and limit sets of Fuchsian groups, *Ergod. Th. and Dynam. Sys.*, 1 (1981), 337-360.

[89] C. Series, Geometrical Markov coding of geodesics on surfaces of constant negative curvature, *Erg. Th. and Dynam. Sys.*, 6 (1986), 601-625.

[90] M. Shub, *Global stability of dynamical systems,* Springer, Berlin, 1989.

[91] Y.G. Sinai, The central limit theorem for geodesic flows on manifolds of constant negative curvature, *Dokl. Akad. Nauk. SSSR,* 133 (1960), 1303-1306.

[92] Y.G. Sinai, The construction of Markov partitions, *Fun. Anal. Appl.,* 2 (1968), 70-80.

[93] Y.G. Sinai, The asymptotic behaviour of the number of closed geodesics on a compact manifold of negative curvature, *Transl. A.M.S.,* 73 (1968), 227-250.

[94] Y.G. Sinai, Gibbs measures in ergodic theory, *Russ. Math. Surv.,* 27 (1972), 21-70.

[95] S. Smale, Differentiable dynamical systems, *Bull. Amer. Math. Soc.,* 73 (1976), 747-817.

[96] D. Sullivan, The density at infinity of a discrete group of hyperbolic motions, *Pub. Math. (IHES),* 50 (1979), 171-209.

[97] D. Sullivan, Discrete conformal groups and measurable dynamics, *Bull. Amer. Math. Soc.*, 6 (1982), 57-73.

[98] T. Sunada, Geodesic flows and random walks, *Adv. Stud. Pure Math.,* 3 (1984), 47-85.

[99] F. Tangerman, Ph.D. thesis, Boston University, 1988.

[100] P. Walters, Ruelle's operator theorem and g-measures, *Trans. Amer. Math. Soc.,* 214 (1975), 375-387.

[101] P. Walters, A variational principle for the pressure of continuous transformations, *Amer. J. Math.,* 97 (1976), 937-971.

[102] P. Walters, *An introduction to ergodic theory,* G.T.M. 79, Springer, Berlin, 1982.

[103] N. Wiener, *The Fourier integral and certain of its applications,* C.U.P., Cambridge, 1967.

W. PARRY
Mathematics Institute
Univ. of Warwick
Coventry, CV 4 7AL
England

M. POLLICOTT
Centro de Matemática
Faculdade de Ciências
Univ. de Porto
4000 Porto, Portugal

RÉSUMÉ

Ce volume a pour sujet principal trois théorèmes qui décrivent, dans leurs contextes appropriés, la distribution dans "l'espace, le temps et la symétrie" d'orbites fermées pour les systèmes hyperboliques. Chacun des résultats est dérivé par des méthodes inspirées de la théorie analytique des nombres, et implique aussi l'analyse d'une fonction générale zéta. La fonction zéta en question est une fonction génératrice pour les orbites fermées et pondérées de la suspension d'un déplacement de type fini (ou d'un flot hyperbolique). Pour déterminer les propriétés analytiques et méromorphiques de cette fonction zéta, on étudie les valeurs caractéristiques d'un opérateur "Ruelle-Perron-Frobenius" associé.

Les chapitres précédents étudient les propriétés de base de déplacements de type fini, le théorème Ruelle-Perron-Frobenius, les états d'équilibre et la pression. De là, on passe aux relations entre les propriétés spectrales de l'opérateur ci-dessus, et aux orbites périodiques d'un flot suspendu. Les méthodes classiques (théorème d'Ikehara) nous permettent ensuite de prouver, dans des conditions modérées, une formule asymptotique pour le nombre d'orbites fermées, et aussi un résultat d'équidistribution spatiale (pondérée) pour les orbites fermées. On prouve aussi un analogue du théorème de Chebotarev, dans le contexte d'extensions de groupe compactes de flots hyperboliques.

Les autres sujets abordés sont

(i) le transfert des résultats d'un contexte symbolique à un contexte de variétés,

(ii) un résultat optimal pour les extensions méromorphiques de la fonction zéta,

(iii) les changements de vélocité et la relation entre la mesure Sinai-Ruelle-

Bowen, et la mesure d'entropie maximale, finalement

(iv) les théorèmes de type Chebotarev pour les extensions $\mathbb{Z}^d$, et la mesure Patterson-Sullivan.

On conclue avec des appendices sur le théorème d'Ikehara, les cocycles unitaires, les partitions de Markov, les flots géodésiques et la théorie des perturbations.