DYNAMICAL SYSTEMS
BANACH CENTER PUBLICATIONS, VOLUME 115
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2018

HYPERBOLIC SYSTEMS, ZETA FUNCTIONS AND OTHER FRIENDS

MARK POLLICOTT

Mathematics Institute, University of Warwick
Coventry, CV4 7AL, United Kingdom
E-mail: masdbl@warwick.ac.uk

Abstract. We discuss a number of inter-related topics, usually ideas from hyperbolic dynamics applied to geometry, fractal geometry, etc. This is based on lectures given at IMPAN, Warsaw.

Contents

2010 *Mathematics Subject Classification*: 37C30, 37F35, 37D20.
Key words and phrases: uniformly hyperbolic systems, dynamical zeta functions.
The paper is in final form and no version of it will be published elsewhere.

DOI: 10.4064/bc115-5

1. Introduction.

A "golden thread" running through these lectures will be dynamical zeta functions, intended to help bind together a number of seemingly disparate topics. In fact, the zeta function can best be viewed as a versatile tool with applications to a wide range of problems.

Having already mentioned dynamical zeta functions, this brings us to a basic question:.

QUESTION. What are zeta functions (in dynamical systems)?

These usually come in two flavours:

1. zeta functions for discrete maps $T : X \to X$; and
2. zeta functions for continuous flows $\phi_t : X \to X$ $(t \in \mathbb{R})$.

As a rough rule of thumb, the zeta function for maps has attracted more attention and has a far greater literature; and the latter is often the more challenging. Let us start from the *discrete* case and return to the *continuous* case later.

1.1. Discrete maps and zeta functions.

Let $T : X \to X$ be a hyperbolic diffeomorphism for a compact manifold. For definiteness, and hopefully clarity, let us consider the specific case of $X = \mathbb{R}^d/\mathbb{Z}^d$, the standard d-dimensional torus. Let $T : X \to X$ be a (linear) hyperbolic toral automorphism, i.e.,

1. let $A \in GL(d, \mathbb{Z})$ with $T(\underline{x} + \mathbb{Z}^d) = A\underline{x} + \mathbb{Z}^d$ (for $\underline{x} \in \mathbb{R}^d$), and
2. the matrix A has no eigenvalues on the unit circle.

Let us recall a very simple and well-known example.

EXAMPLE 1.1 (Arnol'd CAT map [3]). We can let $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ and then define $T : \mathbb{T}^2 \to \mathbb{T}^2$ by $T(x, y) = (2x + y, x + y) \pmod 1$.

Let us return to the definition of the zeta function for T. We denote by

$$\mathrm{Fix}(T^n) = \{\underline{x} \in \mathbb{T}^2 : T^n\underline{x} = \underline{x}\}$$

the set of points on the torus fixed by T^n. For any hyperbolic diffeomorphism it is a standard fact that the set of fixed points of a given period will be finite. Moreover, the hyperbolicity ensures that this number grows at an exponential rate.

The definition of the zeta function in this case is illustrative of the definition in the general case. Following Artin and Mazur we have the following definition of a zeta function [4].

DEFINITION 1.2. The zeta function $\zeta(z)$ associated to a map $T : X \to X$ is a complex function given by

$$\zeta(z) := \exp\left(\sum_{n=1}^{\infty} \frac{z^n}{n} \#(\mathrm{Fix}(T^n))\right)$$

for $z \in \mathbb{C}$.

For the case of hyperbolic toral automorphisms the right hand side converges for $|z|$ sufficiently small. The definition for general Anosov maps $f : M \to M$ is completely analogous. We recall that a diffeomorphism $f : M \to M$ of a compact manifold is *Anosov* if:

1. there is a continuous splitting $TM = E^s \oplus E^u$ and constants $C > 0$ and $0 < \lambda < 1$ such that $\|Df^n|E^s\| \leq C\lambda^n$ and $\|Df^{-n}|E^u\| \leq C\lambda^n$, for $n \geq 0$;
2. $f : M \to M$ is transitive, i.e., there exists a dense orbit.

All of this leads to the following natural questions.

Question. Can we extend $\zeta(z)$ to a larger domain in z? Where are the zeros and poles (or singularities) for this extension?

For this particular case of orientation preserving hyperbolic total automorphisms, the answers to these two questions are relatively easy [65].

Theorem 1.3. *For an orientation preserving hyperbolic toral automorphism the zeta function $\zeta(z)$ extends to $\mathbb{C}$ (as a rational function $p(z)/q(z)$ with $p, q \in \mathbb{R}[z]$).*

Fortunately, in this case the proof of the result is very simple. In particular, this is a special case of the famous Lefschetz fixed point theorem., i.e., since $\det A = 1$ we have

$$\#(\mathrm{Fix}(T^n)) = \sum_{k=0}^{d} (-1)^{k+1} \mathrm{tr}(T_*^n : H_k \to H_k)$$

where $T_* : H_k \to H_k$ is the induced linear map on the kth real homology group, as observed by Smale [65]. The key point here is that the toral automorphism is assumed to be orientation preserving and thus the Lefschetz index for each fixed point is 1. Let us consider the specific example of the Arnol'd CAT map again.

Example 1.4. We can let $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ and then $\mathrm{tr}(A^n) - 2 = \#(\mathrm{Fix}(T^n))$. A simple computation gives

$$\zeta(z) = \exp\left(\sum_{n=1}^{\infty} \frac{z^n}{n}(\mathrm{tr}(A^n) - 2)\right) = \frac{(1-z)^2}{\det(I - zA)}.$$

More generally, the zeta function has a rational extension to $\mathbb{C}$ for any Anosov diffeomorphism, or more generally Axiom A diffeomorphisms as was originally proved by A. Manning [37]. The smallest pole (in terms of its absolute value) comes from the radius of convergence of the series:

$$1/R = \lim_{n \to +\infty} \#(\mathrm{Fix}(T^n))^{1/n} =: \lambda$$

where λ is the maximal eigenvalue of the matrix A. In particular, $\log \lambda$ is the topological entropy $h(T)$ of $T : \mathbb{T}^d \to \mathbb{T}^d$. The other zeros and poles of $\zeta(z)$ reflect the speed of convergence in this limit.

1.2. Continuous flows and zeta functions. Let us next turn to the case of flows. But first let us recall a (more) famous zeta function from number theory defined in terms of the prime numbers $p = 2, 3, 5, 7, 11, \ldots$.

DEFINITION 1.5 (Riemann zeta function [20]). We define the *Riemann zeta function* by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p (1 - p^{-s})^{-1}, \qquad s \in \mathbb{C},$$

where the product is over all prime numbers.

The equivalence of the two definitions comes from the simple expansion

$$\frac{1}{1 - p^{-s}} = 1 + p^{-s} + p^{-2s} + p^{-3s} + \cdots$$

for $Re(s) > 1$. This converges for $Re(s) > 1$ to a nonzero analytic function. The following results are classical in number theory:

1. $\zeta(s)$ has a meromorphic extension to $\mathbb{C}$; and
2. the zeros for $\zeta(s)$ are mysterious (e.g., the Riemann Hypothesis remains open, stating that the zeros in the critical strip $0 < Re(s) < 1$ lie on the line $Re(s) = \frac{1}{2}$).

Returning to the definition of zeta functions for flows, we can consider a simple example which illustrates how things work, before giving the definition in the general case.

EXAMPLE 1.6 (Suspension flow). Consider the simple setting of a Cantor set X and the classical Smale horseshoe map $T : X \to X$ [65]. This is a diffeomorphism of the sphere S^2 which maps a rectangle on S^2 across itself in a horseshoe shape. Then the Cantor set X corresponds to points whose entire orbit is contained in the rectangle and it is homeomorphic to the sequence space $\Sigma = \{0,1\}^{\mathbb{Z}} = \{x = (x_n) : x_n \in \{0,1\}\}$ and T is conjugate to the shift map $\sigma : \Sigma \to \Sigma$. We may introduce a function $r : X \to \mathbb{R}^+$ that depends only on the zeroth coordinate x_0 and is defined by

$$r(x) = \begin{cases} \alpha & \text{if } x_0 = 0, \\ \beta & \text{if } x_0 = 1, \end{cases}$$

where $0 < \alpha < \beta$ [47].

We can then define by

$$\Lambda^r = \{(x,u) : 0 \leq u \leq r(x)\}/(x, r(x)) \sim (Tx, 0)$$

the area under the graph of r, where the points $(x, r(x)))$ and $(Tx, 0)$ are identified. We then define $\phi_t : \Lambda^r \to \Lambda^r$ by $\phi_t(x,u) = (x, u+t)$ subject to the identifications. There is then a natural bijection between closed orbits for $T : \Lambda \to \Lambda$ and $\phi_t : \Lambda^r \to \Lambda^r$ such that $\{x, Tx, \ldots, T^{n-1}x\}$ corresponds to a closed orbit τ of period

$$\lambda(\tau) = r(x) + r(Tx) + \cdots + r(T^{n-1}x).$$

We can now define a zeta function for the flow (in the example above).

DEFINITION 1.7 ([59]). We can formally define a zeta function for ϕ by

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

where τ is a prime periodic orbit for ϕ (i.e., not a multiple of a periodic orbit of shorter period). This converges for $Re(s)$ sufficiently large.

More generally, we can similarly define the zeta function for Axiom A flows.

As we can see, the zeta function for flows is defined by analogy with the Euler product form of the Riemann zeta function $\zeta(s)$, where the primes are replaced by the exponentials of the least periods of orbits.

In the present context, the following is a simple exercise.

LEMMA 1.8. *For the example above we can write*

$$\zeta_\phi(s) = \exp\Big(\sum_{n=1}^\infty \frac{1}{n} \sum_{T^n x = x} e^{-sr^n(x)}\Big)$$

where $T^n x = x$ is a fixed point for x.

Proof. Providing $Re(s)$ is sufficiently large, we can write

$$\begin{aligned}\prod_\tau (1 - e^{-s\lambda(\tau)})^{-1} &= \exp\Big(-\sum_\tau \log(1 - e^{-s\lambda(\tau)})\Big) = \exp\Big(\sum_{m=1}^\infty \sum_\tau \frac{e^{-sm\lambda(\tau)}}{m}\Big)\\ &= \exp\Big(\sum_{n=1}^\infty \sum_{\{x,\dots,T^{n-1}x\}(\text{prime})} \sum_{m=1}^\infty \frac{1}{m} e^{-smr^n(x)}\Big)\\ &= \exp\Big(\sum_{n=1}^\infty \sum_{T^n x = x(\text{prime})} \frac{1}{n} \sum_{m=1}^\infty \frac{e^{-smr^n(x)}}{m}\Big) = \exp\Big(\sum_{l=1}^\infty \frac{1}{l} \sum_{T^l x = x} e^{-sr^l(x)}\Big)\end{aligned}$$

which completes the proof. ■

In the particular case that the roof function is constant (i.e., $\alpha = \beta$) the dynamical zeta function for the flow in this example can be written in terms of the zeta function for the discrete map.

REMARK 1.9. If $\alpha = \beta$ then $\zeta_\phi(s) = 1/(1 - 2e^{-s\alpha})$ (i.e., the continuous zeta function is related to the discrete zeta function with $z = e^{-s\alpha}$). If $\alpha \neq \beta$ then can write

$$\lambda(\tau) = \alpha \text{Card}\{0 \le j \le n-1 : x_j = 0\} + \beta \text{Card}\{0 \le j \le n-1 : x_j = 1\}$$

and then we have

$$\zeta_\phi(s) = \exp\Big(\sum_{n=1}^\infty \frac{1}{n}(e^{-s\alpha} + e^{-s\beta})^n\Big) = \frac{1}{1 - e^{-s\alpha} - e^{-s\beta}}$$

[60], [49]. Thus if $h > 0$ is a unique solution to $e^{-h\alpha} + e^{-h\beta} = 1$ then:

1. for $Re(s) > h$ we have that $\zeta_\phi(s)$ converges to a nonzero analytic function;
2. h is a simple pole for $\zeta_\phi(s)$;
3. $\zeta_\phi(s)$ has a meromorphic extension to $\mathbb{C}$; and
4. if α/β is irrational then there are poles $s_n = \sigma_n + it_n$ satisfying $1 = e^{-s_n\alpha} + e^{-s_n\alpha}$ for which $\sigma_n \nearrow h$. This follows from properties of almost periodic functions.

The value h can be shown to be the topological entropy for the associated flow (i.e., the topological entropy of the time one flow $\phi_{t=1}$).

In the next section we will begin to show that these dynamical zeta functions have practical applications to apparently unrelated problems.

2. Dynamically defined Cantor set and averaging transfer operators. We begin with an application to the Hausdorff dimension of limit sets for iterated function schemes [25].

Let $X \subset [0,1]$ be a dynamically defined Cantor set. More precisely, let $T_0, T_1 : [0,1] \to [0,1]$ be C^ω (or more generally C^1) contractions with disjoint images (i.e., $T_0[0,1] \cap T_1[0,1] = \emptyset$). The associated Cantor set X is the unique nonempty closed set $X \subset [0,1]$ such that

$$T_0 X \cup T_1 X = X.$$

We can define a locally distance expanding map $T : X \to X$ by

$$T(x) = \begin{cases} T_0^{-1}(x) & \text{if } x \in TX_0, \\ T_1^{-1}(x) & \text{if } x \in TX_1. \end{cases}$$

We recall some classical examples.

Example 2.1 (Middle 1/3-Cantor set). Let $T_0(x) = x/3$ and $T_1(x) = x/3 + 2/3$. Then

$$X = \left\{ x = \sum_{n=1}^{\infty} \frac{x_n}{3^{n+1}} : x_n \in \{0,2\} \right\}$$

(i.e., a triadic expansion with coefficients either 0 or 2). We can define $T : X \to X$ by $T(x) = 3x \pmod 1$.

The next example is similar, but defined using nonlinear contractions.

Example 2.2 (E_2). Let $T_0(x) = \frac{1}{1+x}$ and $T_1(x) = \frac{1}{2+x}$. Then

$$X = \{x = [a_1, a_2, a_3, \dots] : a_n \in \{1,2\}\}$$

i.e., the points whose continued fraction expansion contains only the digits 1 and 2. We can define the expanding map $T : X \to X$ by $Tx = 1/x - [1/x]$.

We would like to quantify the size of these Cantor sets. The natural notion is the Hausdorff dimension (although for these examples the Hausdorff dimension coincides with the more easily defined box dimension).

Question. What is the Hausdorff dimension of the Cantor sets X in these examples?

In particular, we need to find some useful way to characterize the dimension. Let $C(X)$ be the space of continuous functions $w : X \to \mathbb{C}$.

Definition 2.3. We define a transfer operator $\mathcal{L} : C(X) \to C(X)$ by

$$\mathcal{L}w(x) = |T_0'(x)| w(T_0 x) + |T_1'(x)| w(T_1 x).$$

Unfortunately, the spectrum of $\mathcal{L} : C(X) \to C(X)$ is rather lacking in fine structure, as the next lemma reveals.

Lemma 2.4. *The spectrum of $\mathcal{L} : C(X) \to C(X)$ is a closed ball whose radius is the norm $\|\mathcal{L}\| = \sup\{\|\mathcal{L}f\|_\infty : \|f\|_\infty \le 1\}$ of the operator (or equivalently the spectral radius of the operator).*

Recall that the spectrum of $\mathcal{L}$ is defined to be the subset of the complex plane:

$$\mathrm{Spec}(\mathcal{L}) = \{z \in \mathbb{C} : (zI - \mathcal{L}) : C(X) \to C(X) \text{ is not invertible}\}.$$

We can illustrate the proof of the above lemma with the first example (Example 2.1), the general case being similar. We first observe that $\|\mathcal{L}\| \leq \frac{2}{3}$ from which we deduce that the spectral radius is at most $\frac{2}{3}$. Fix $w_0 \in C(X)$ such that $\mathcal{L}w_0(x) = 0$ for all $x \in X$ (e.g., $w_0(x) = 1 - w_0(1-x)$). For any $|\lambda| < 1$ we can define

$$w_\lambda(x) := \sum_{n=0}^{\infty} \lambda^n w_0(T^n x) \in C(X)$$

since $C(X)$ is a Banach space. But $\frac{2}{3}\lambda$ is an eigenvalue, since

$$\mathcal{L}w_\lambda(x) = \underbrace{\mathcal{L}w_0(x)}_{=0} + \sum_{n=1}^{\infty} \lambda^n \mathcal{L}(w_0 \circ T^n)(x) = \sum_{n=1}^{\infty} \lambda^n (w_0 \circ T^{n-1})(x) = \frac{2}{3}\lambda w_\lambda(x)$$

and $\mathcal{L}(w_0 \circ T^n)(x) = \frac{2}{3}(w_0 \circ T^{n-1})(x)$, unless $w_\lambda = 0$, which we can assume, without loss of generality, is not the case. This completes the proof of the lemma.

To further our understanding of the zeta function, we want to consider transfer operators with smaller spectra. In particular, we need Banach spaces with "fewer" functions for the transfer operators to act upon, an issue which we will address in the next section. Moreover, to add more utility to these operators we would like to change the weights to include a parameter $s \in \mathbb{R}$ (or even $s \in \mathbb{C}$).

DEFINITION 2.5. Given $s \in \mathbb{R}$ ($s \in \mathbb{C}$) we can define a family of operators $\mathcal{L}_s : C(X) \to C(X)$ by

$$\mathcal{L}_s w(x) = |T_0'(x)|^s w(T_0 x) + |T_1'(x)|^s w(T_1 x).$$

More generally, we could consider a finite family of contractions $T_1, \ldots, T_n$ and define the operators $\mathcal{L}_s : C(X) \to C(X)$ by

$$\mathcal{L}_s w(x) = \sum_{j=1}^{n} |T_j'(x)|^s w(T_j x).$$

We can illustrate the transfer operator using our two previous examples.

EXAMPLE 2.6. 1. For the middle $\frac{1}{3}$-Cantor set we have a transfer operator

$$\mathcal{L}_s w(x) = \left(\frac{1}{3}\right)^s w\left(\frac{x}{3}\right) + \left(\frac{1}{3}\right)^s w\left(\frac{x+3}{3}\right).$$

2. For E_2 we have a transfer operator

$$\mathcal{L}_s w(x) = \left(\frac{1}{x+1}\right)^{2s} w\left(\frac{1}{x+1}\right) + \left(\frac{1}{x+2}\right)^{2s} w\left(\frac{1}{x+2}\right).$$

The next step is to find a suitable Banach space $B \subset C(X)$ for which the operator $\mathcal{L} : B \to B$ has better spectral properties and then use these to deduce interesting results about X and $T : X \to X$.

3. Banach spaces of analytic functions. There are many candidates for spaces of functions upon which we can act with the transfer operator. Perhaps the simplest principle is to consider the smallest space preserved by the transfer operator associated to the transformation T. For the present, we will consider those T which are analytic (as in the two examples above) and Banach spaces of analytic functions since these are preserved by the transfer operator

Let U be an open ball in $\mathbb{C}$. Let $B = B(U)$ be the Banach space of bounded analytic functions $w : U \to \mathbb{C}$ with the norm

$$\|w\| = \|w\|_\infty := \sup_{z \in U} |w(z)|.$$

(The completeness comes from Montel's Theorem in complex analysis.)

The advantage of transfer operators that preserve Banach spaces of analytic functions is that they take a special form, which we will now describe.

Definition 3.1. We say that a bounded linear operator $T : B \to B$ is nuclear (or trace class) if we can write

$$T(\cdot) = \sum_{n=0}^{\infty} \lambda_n l_n(\cdot) w_n$$

where

1. $w_n \in B$ with $\|w_n\| = 1$;
2. $l_n \in B^*$ with $\|l_n\| = 1$; and
3. $|\lambda_n| = O(\theta^n)$, for some $0 < \theta < 1$.[1]

Remark 3.2. Nuclear operators are automatically compact operators, as is easily seen from the definition, and thus only have countably many isolated eigenvalues all of which, except the one at zero, are isolated.

In the context of dynamically defined Cantor sets, let $T_i : [0,1] \to [0,1]$ $(i = 1,2)$ be analytic and assume there are nested open sets

$$[0,1] \subset U \subset U^+ \subset \mathbb{C}$$

in the complex plane such that the maps extend analytically to U^+ and satisfy

$$\text{closure}(T_i U^+) \subset U.$$

By looking at the spectrum of the operators on the smaller space of analytic functions we see that the spectrum of the operator has much more structure, which ultimately gives us more information about, for example, the zeta function. The most useful result in this direction is the following [58].

Theorem 3.3 (Grothendieck-Ruelle). *The operators $\mathcal{L}_s : B \to B$ $(s \in \mathbb{C})$ are nuclear.*

Rather than discussing the implications of this theorem in complete generality, let us consider specific cases. These are best illustrated by considering the previous two examples.

Example 3.4 (Middle 1/3-Cantor set). Let us choose

$$U = \left\{ z \in \mathbb{C} : |z| < \frac{5}{2} \right\} \quad \text{and} \quad U^+ = \{ z \in \mathbb{C} : |z| < 3 \},$$

[1] This is slightly stronger than the usual definition of a nuclear operator, but is sufficient for our purposes.

say. Then a simple calculation shows

$$T_0(U^+) = \left\{ z \in \mathbb{C} : |z| < \frac{3}{2} \right\} \subset U \text{ and } \quad \text{and} \quad T_1(U^+) = \left\{ z \in \mathbb{C} : \left| z - \frac{1}{2} \right| < \frac{3}{2} \right\} \subset U.$$

In particular, $\mathcal{L}_s(B(U)) \subset B(U^+)$. Such operators are referred to as "analyticity improving" since functions in the image are analytic on a larger domain than they initially were. By Cauchy's theorem (which can be applied by virtue of $\partial U \subset U^+$) we can write

$$\mathcal{L}_s w(z) = \frac{1}{2\pi i} \int_{|\xi|=5/2} \frac{\mathcal{L}_s w(\xi)}{z - \xi} d\xi = \sum_{n=0}^{\infty} \lambda_n w_n(z) l_n(w)$$

for $z \in U^+$ where:

(a) $w_n(z) = z^n \in B$; and
(b) $l_n(w) \asymp \frac{1}{2\pi i} \int_{|\xi|=5/2} \frac{\mathcal{L}_s w(\xi)}{\xi^{n+1}} d\xi$

where $\|l_n\| = 1$ and

$$\lambda_n = \left| \frac{1}{2\pi i} \int_{|\xi|=5/2} \frac{\mathcal{L}_s w(\xi)}{\xi^{n+1}} d\xi \right|.$$

It is easy to see that $\lambda_n = O(\theta^n)$ with $\theta = \frac{5}{6}$.

The case of the nonlinear Cantor set is slightly more interesting.

Example 3.5 (E_2). Let us choose

$$U = \left\{ z \in \mathbb{C} : |z - 1| < \frac{3}{2} \right\} \quad \text{and} \quad U^+ = \left\{ z \in \mathbb{C} : |z - 1| < \frac{19}{12} \right\},$$

say. Then a simple (although not quite as simple as in the previous example) calculation gives

$$T_0 U^+ = \left\{ z \in \mathbb{C} : \left| z - \frac{288}{215} \right| < \frac{228}{215} \right\} \subset U \quad \text{and} \quad T_1 U^+ = \left\{ z \in \mathbb{C} : \left| z - \frac{432}{935} \right| < \frac{228}{935} \right\} \subset U.$$

By Cauchy's theorem (since $\partial U \subset U^+$) we can write

$$\mathcal{L}_s w(z) = \frac{1}{2\pi i} \int_{|\xi-1|=3/2} \frac{\mathcal{L}_s w(\xi)}{z - \xi} d\xi = \sum_{n=0}^{\infty} \lambda_n w_n(z) l_n(w)$$

where

(a) $w_n(z) = (z-1)^n \in B$;
(b) $l_n(w) \asymp \frac{1}{2\pi i} \int_{|\xi-1|=3/2} \frac{\mathcal{L}_s w(\xi)}{\xi^{n+1}} d\xi$

where $\|l_n\| = 1$ and

$$\lambda_n = \left| \frac{1}{2\pi i} \int_{|\xi-1|=3/2} \frac{\mathcal{L}_s w(\xi)}{\xi^{n+1}} d\xi \right|.$$

It is easy to see that $\lambda_n = O(\theta^n)$ with $\theta = \frac{18}{19}$.

Now that we have introduced a suitable Banach space of analytic functions for the transfer operators to act upon, it still remains to relate these to the zeta functions we previously defined. There are three useful facts (which we will elaborate upon later) that we list below for our immediate convenience:

Properties of the operators $\mathcal{L}_s$ acting on analytic functions. The following properties will be useful (see [29], [58], [33]).

1. The operators $\mathcal{L}_s : B \to B$ are nuclear and so we can define a function of two variables $(z, s \in \mathbb{C})$
$$d(z,s) := \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \mathrm{trace}(\mathcal{L}_s^n)\right)$$
(which converges for $|z|$ sufficiently small, depending on $Re(s)$).
2. We can explicitly compute
$$\mathrm{trace}(\mathcal{L}_s^n) = \sum_{T^n x = x} \frac{|(T^n)'(x)|^s}{1 - (T^n)'(x)}.$$
3. $d(z, s)$ has an analytic extension to $\mathbb{C}^2$. Moreover, we can expand
$$d(z,s) = 1 + \sum_{n=1}^{\infty} a_n(s) z^n$$
where there exists $C > 0$ such that $|a_n(s)| \leq C\theta^{n^2}$, with explicit expressions for $a_n(s)$ in terms of $(T^m)'(x)$, where $T^m x = x$, $m \leq n$.

This has an immediate application to zeta functions.

Proposition 3.6. *We can write $\zeta_\phi(s) = d(1, s+1)/d(s)$ with $r = -\log|T'|$ to give the connection with the zeta function $\zeta_\phi(s)$.*

The Cantor set E_2 can be generalized to those points whose continued fraction expansions are uniformly bounded. This links nicely to the following classical open problem:

Remark 3.7 (Zaremba Conjecture (1971)). There exists $N \in \mathbb{N}$ such that
$$\left\{q \in \mathbb{N} : \frac{p}{q} = [a_1, a_2, \ldots, a_N] \text{ for } a_i \in \{1,2,3,4,5\}\right\} = \mathbb{N}.$$
Bourgain and Kontorovich proved the set on the left hand side has density 1 [11], [36].

There are also classical questions and results on the differences of linear Cantor sets. In the context of a nonlinear Cantor set (coming from bounded continued fraction expansions) we mention the following nice result.

Remark 3.8 (C. Moreira [45]). The difference set $E_2 - E_2$ has full dimension, i.e., $\dim_H(E_2 - E_2) = 1$.

4. Applications of zeta functions.

We will return to discussing the properties of the zeta functions after considering some applications.

4.1. Application I: Computing Hausdorff dimension.

For definiteness, let us again consider the nonlinear Cantor set X $(= E_2)$ with continued fraction coefficients 1 or 2. Unlike the case of linear Cantor sets, there is no simple formula for the dimension of the limit set. However, there is an expression which doesn't (at first sight) seem particularly useful [33].

LEMMA 4.1. *The real number $s = \dim_H(X)$ is a zero for*

$$d(1,s) = \exp\Big(-\sum_{n=1}^{\infty}\frac{1}{n}\sum_{T^n x = x}\frac{|(T^n)'(x)|^s}{1-(T^n)'(x)}\Big)$$

where the sum over periodic points corresponds to numbers with periodic continued fraction expansions.

Proof. This follows from Bowen's formula [12], [61] characterizing $\dim_H(X)$ as the zero of a function $P(s)$ defined in terms of the maximal eigenvalue of the transfer operator (and called the pressure). In fact, the (first) zero appears at the value $s \in \mathbb{R}$ where

$$e^{P(s)} := \lim_{n\to+\infty}\Big(\sum_{T^n x = x}\frac{|(T^n)'(x)|^s}{1-(T^n)'(x)}\Big)^{1/n} = \lim_{n\to+\infty}\Big(\sum_{T^n x = x}|(T^n)'(x)|^s\Big)^{1/n} = 1. \blacksquare$$

The first encouraging sign is that the fixed points are simply quadratic surds (i.e., algebraic numbers of degree two). However, more importantly there is an expansion of $d(1,s)$ in terms of a rapidly converging series. Writing

$$d(1,s) = 1 + \sum_{n=1}^{\infty} a_n(s)$$

where $|a_n(s)| = O(\theta^{n^2})$, $\theta = (4/5)^{1/4})$ we can approximate $d(1,s)$ by the polynomial

$$d_N(1,s) = 1 + \sum_{n=1}^{N} a_n(s)$$

and then s_N satisfies $d_N(1,s_N) = 0$ with $s_N = \dim_H(X) + O(\theta^{N^2})$.

Using a more elaborate variant of this approach we have the following result [34]:

THEOREM 4.2 (Jenkinson-Pollicott). *We can write*

$$\begin{aligned}\dim_H(E_2) = 0.&531280506277205141624468647368471785493059109018398\\ &779888397803927529535643831345918109570181185239 8\ldots\end{aligned}$$

accurate to 100 decimal places.

The proof involves choosing $N = 25$. This value of N is sufficiently small to allow a computer assisted numerical computation of $d_N(1,s)$ and yet large enough that the difference between $d_N(1,s)$ and $d(1,s)$ is sufficiently small that their zeros are close. In particular the zero of $d_N(1,s)$ can be easily estimated to a high degree of accuracy, using a delicate combination of numerical and theoretical bounds. This leads to an approximation of the zero of $d(1,s)$, i.e., the Hausdorff dimension $\dim_H(E_2)$.

4.2. Application II: Selberg zeta function. The original application of transfer operators to the theory of zeta functions associated to geodesics on (Riemann) surfaces dates back to Ruelle's original paper [58] (see also [53]). To illustrate the basic ideas, we will consider the partially simple example of a pair of pants V, which is a Riemann surface of constant curvature $\kappa = -1$ with infinite area arising from three infinite funnels. We can write $V = \mathbb{H}^2/\Gamma$ where $\mathbb{H}^2 = \{z = x + iy : y > 0\}$ denotes the upper half plane with the Poincaré metric $ds^2 = (dx^2 + dy^2)/y^2$ and $\Gamma = \langle R_1, R_2, R_3\rangle$ is the free group generated by certain isometries $R_1, R_2, R_3 : \mathbb{H}^2 \to \mathbb{H}^2$. We first describe this construction in a little more detail.

Example 4.3 (A pair of pants). Let $[a_0, b_0], [a_1, b_1], [a_2, b_2] \subset \mathbb{R}$ be disjoint intervals in the real line, with centres $c_j = (a_j + b_j)/2$ and $r_j = (b_j - a_j)/2$ for $j = 1, 2, 3$. Let $R_j : \mathbb{R} \cup \{\infty\} \to \mathbb{R} \cup \{\infty\}$ be the linear fractional transformation defined by

$$R_j(x) = \frac{r_j^2}{x - c_j} + c_j,$$

for $j = 1, 2, 3$. This extends to the upper half plane $\mathbb{H}^2$ by

$$R_j(z) = r_j^2 \frac{\overline{z} - \overline{c_j}}{|z - c_j|^2} + c_j,$$

for $j = 1, 2, 3$. To construct the appropriate Banach space of analytic functions, we choose disjoint (larger) disks

$$D_j = \{z \in \mathbb{C} : |z - c_j| < t_j\} \supset [a_j, b_j]$$

for suitable radii $t_j > r_j$, for $j = 1, 2, 3$. For $j \neq l$ we arrange the radii such that closure$(R_l(D_j)) \subset D_l$.

By analogy with the Banach spaces of analytic functions introduced to deal with the Hausdorff dimension of dynamically defined Cantor sets, we can consider analytic functions on the disks D_1, D_2 and D_3. More precisely, let $B = B(\bigcup_{j=1}^3 D_j)$ denote bounded analytic functions on the union $\bigcup_{j=1}^3 D_j$ of disjoint disks and then $\mathcal{L}_s : B \to B$ is defined by

$$\mathcal{L}_s w(z) = \sum_{j \neq l} |R_j'(z)|^s w(R_j z) \qquad \text{for } z \in D_l.$$

We can now write the associated zeta function as

$$d(s) = Z(s) := \prod_{\gamma} \prod_{n=0}^{\infty} (1 - e^{-(s+n)l(\gamma)}) \tag{4.1}$$

where γ is a primitive closed geodesic on the pair of pants V of length $l(\gamma)$. The quotient surface V is an infinite volume surface of curvature $\kappa = -1$.

Remark 4.4. The limit set of $\Gamma = \langle R_1, R_2, R_3 \rangle$ is the Cantor set of accumulation points (in the Euclidean sense) of the orbit Γi of $i \in \mathbb{H}^2$. It is a nonlinear Cantor set of Hausdorff dimension $\delta = \dim_H(X)$.

Remark 4.5. The recurrent part of the geodesic flow is coded by sequences and the transition matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

This is a very simplified form of the Bowen-Series coding used to code geodesics on convex co-compact surfaces[15], [64]. The coding can be naturally realized in terms of the limit set, and the roof function on the limit set takes the form $r(x) = \log |R_j'(x)|$ for $x \in D_j$.

We conclude from the properties of the determinant $d(z, s)$ the following result:

Theorem 4.6. *The zeta function $Z(s)$ extends analytically to the entire complex plane $\mathbb{C}$.*

The classical approach to studying zeta functions on finite area surfaces V of curvature $\kappa = -1$ uses the Selberg trace formula and unitary representations in the Hilbert space $L^2(V)$. However, in the case of infinite area surfaces this is less natural and the dynamical approach to the zeta function $Z(s)$ is essentially the only approach available to extending the zeta function.

Remark 4.7. The largest zero appears at $\delta = \lambda(1-\lambda)$ where $\lambda > 0$ is the smallest eigenvalue of the Laplacian. The other zeros for $Z(s)$ in some special cases cases were plotted by Borthwick [9], where the zeros appear to be described in terms of specific curves. An explanation of this appears in [55].

4.3. Application III: Circle packings. In the previous section we considered a Fuchsian group Γ whose limit set is a Cantor set in the real line $\mathbb{R}$. In this section we consider a higher dimensional analogue where the Fuchsian group is replaced by a Kleinian group and the limit set is now in $\mathbb{C}$, called the Apollonian circle packing $\mathcal{C}$. This is the closure of a countable union of closed circles. Moreover, the radii r_n of the circles satisfy $r_n \to 0$ as $n \to +\infty$.

Let $\delta = \dim_H(\mathcal{C})$ denote the Hausdorff dimension of the set $\mathcal{C}$. We have the following simple counting result for the radii of the circles [38].

Theorem 4.8 (Kontorovich-Oh, 2009). *There exists $C > 0$ such that*

$$\#\{r_n \geq \epsilon\} \sim C\epsilon^{-\delta}$$

as $\epsilon \to 0$ (i.e., $\lim_{\epsilon\to 0} \epsilon^{\delta}\mathrm{Card}\{r_n \geq \epsilon\} = C$).

We want to describe an alternative viewpoint of this theorem, contained in [54].

Step 1. Let C_1, C_2, C_3, C_4 be four initial mutually tangent circles in $\mathcal{C}$.

Step 2. Following a result of Beecroft from 1842, let K_1, K_2, K_3, K_4 be the four dual circles (i.e., the circles passing through triples of points chosen from the four tangent points).

Step 3. To introduce the dynamical perspective, let T_1, T_2, T_3, T_4 be reflections in the four circles K_1, K_2, K_3, K_4.

Step 4. All the circles in $\mathcal{C}$ are generated by reflecting C_1, C_2, C_3, C_4 repeatedly under T_1, T_2, T_3, T_4. Consider one of the four curved triangles X coming from the original four tangent circles.

Step 5. Following an approach of Mauldin-Urbański [43] we can generate the circles using the uniformly contractive maps $\phi_{\underline{i}} = f_i \circ f_j^n : X \to X$, with $\underline{i} = (i,j)$ for $i, j = 1, 2, 3, 4$ with $i \neq j$ and $n \geq 1$, where $f_l = T_4 \circ T_l$ for $l = 1, 2, 3$. In particular, by taking the images of the central circle K_4 under iterates of the maps ϕ_i.

Finally, to get the asymptotic formula in the theorem, we want to consider the complex function

$$\eta(s) := \sum_{n=1}^{\infty} r_n^s = \int_1^{\infty} t^{-s} d\pi(t)$$

where $\pi(t) = \mathrm{Card}\{r_n \geq 1/t\}$ is a monotone increasing function and the integral above is understood as a Riemann-Stieltjes integral. For fixed z_0 we can "replace" (or approximate)

$\{r_n\}$ by the derivatives $\{(\phi_{\underline{i}_1} \circ \cdots \circ \phi_{\underline{i}_m})'(z_0)\}$ and replace $\eta(s)$ by

$$\eta_0(s) = \sum_{n=1}^{\infty} \mathcal{L}_s^n \rho(z_0)$$

where $\mathcal{L}_s w(z) = \sum_\phi |\phi'(z)|^s w(\phi z)$ and

$$\rho(z) = \sum_{l=0}^{\infty} |(f_i^l)'(z)|^s.$$

The connection between the domain of $\eta(s)$ and the asymptotic formulae comes from classical Tauberian theorems. Before describing these let us consider a simplified situation.

Remark 4.9 (Motivation for Tauberian theorems). Recall that for Anosov diffeomorphisms,

$$\zeta(z) = \exp\left(\sum_{n=1}^{\infty} \frac{z^n}{n} \text{Card Fix}(T^n)\right) = \frac{P(z)}{Q(z)},$$

a rational function. For example for the hyperbolic toral automorphism (in Example 1.4) we can write

$$\zeta(z) = \frac{(1-z)^2}{\det(I - zA)}.$$

Therefore, denoting by $\lambda = e^{h(T)}$ the maximum eigenvalue of the matrix A, we have

$$\frac{\partial}{\partial z} \log \zeta(z) = \sum_{n=1}^{\infty} z^{n-1} \text{Card Fix}(T^n) = \frac{\lambda}{1 - z\lambda} + \Phi(z)$$

where $\Phi(z)$ is a rational function with poles and zeros in $|z| > R$. We can also write

$$\frac{\lambda}{1 - z\lambda} = \sum_{n=0}^{\infty} z^n \lambda^{n+1}.$$

Thus

$$\sum_{n=1}^{\infty} z^{n-1} (\lambda^{n+1} - \text{Card Fix}(T^n))$$

is analytic in a neighbourhood of $|z| \leq R$. In particular, we deduce that

$$\text{Card Fix}(T^n) = \lambda^n + O(1/R^n)$$

as $n \to +\infty$.

For flows the situation is a little more complicated, but in the same spirit. For flows we would write a Stieltjes integral:

$$\eta(s) = \int_0^{\infty} t^{-s} d\pi(t).$$

The next result provides the appropriate Tauberian machinery required to translate analyticity results on $\eta(s)$ into an asymptotic result [24].

Lemma 4.10 (Tauberian theorem). *If $\eta(s)$ has an analytic extension to a neighbourhood of $Re(s) \geq h$, except for a simple pole of the form $\frac{1}{s-h}$, then $\lim_{t \to +\infty} \frac{\pi(t)}{e^{ht}} = 1$ (i.e., $\pi(t) \sim e^{ht}$).*

Theorem 6.1 now follows from the Ikehara Wiener Tauberian theorem. In particular, we can show that

1. $\eta(s)$ is analytic for $Re(s) > \delta$;
2. $\eta(s)$ has a simple pole at $s = \delta$, with residue $C > 0$;
3. $\eta(s)$ has no poles $s = \delta + it$ where $t \neq 0$.

We can then deduce from the Ikehara Tauberian theorem (Theorem 6.1) that

$$\pi(t) \sim Ct^{\delta} \quad \text{as } t \to +\infty.$$

5. Properties of the transfer operator. Returning to properties of the operators $\mathcal{L}_s$, we first want to explain how the functions $d(z,s)$ can be expressed in terms of periodic points. Key to this is recalling that $\mathcal{L}_s$ is nuclear (or trace class) and the following result.

LEMMA 5.1. *Let $T : X \to X$ be the expanding C^{ω} map. We can write*

$$\operatorname{tr}(\mathcal{L}_s^n) = \sum_{T^n x = x} \frac{|(T^n)'(x)|^s}{1 - ((T^n)'(x))^{-1}}.$$

Proof. We will follow the method used in [44]. We will consider the case $n = 1$, the other cases being similar. Let $T_j x_j = x_j$ be fixed points of contractions $T_j : X \to X$ (and thus fixed points of $T : X \to X$). We can then use the linearity of the trace to write

$$\operatorname{tr}(\mathcal{L}_s) = \sum_j \operatorname{tr}(\mathcal{L}_{s,j})$$

where each of the operators

$$\mathcal{L}_{s,j} w(x) = w(T_j x)|T_j'(x)|^s.$$

is also nuclear. For each j consider the eigenvalue equation

$$\mathcal{L}_{s,j} w(x) = \lambda w(x)$$

with eigenvalue λ and evaluate at $x = x_j$. If $w(x_j) \neq 0$ then $\lambda = |T_j'(x_j)|^s$. If $w(x_j) = 0$ then differentiate again:

$$w'(T_j x)T_j'(x)|T_j'(x)|^s + w'(T_j x)T_j'(x)\frac{\partial}{\partial x}|T_j'(x)|^s = \lambda w'(x).$$

We can evaluate this at

$$w'(x_j)T_j'(x_j)|T_j'(x_j)|^s = \lambda w'(x_j).$$

If $w'(x_j) \neq 0$ then $\lambda = T_j'(x)|T_j'(x)|^s$, etc. Proceeding inductively, for each $k \geq 0$,

$$\lambda = (T_j'(x))^k |T_j'(x)|^s$$

is an eigenvalue for $\mathcal{L}_{s,j}$. Then by summing over $k \geq 0$ we have the trace

$$\operatorname{tr}(\mathcal{L}_{s,j}) = \Big(\sum_{n=1}^{\infty} (T_j'(x_j))^k\Big)|T_j'(x_j)|^s = \frac{|T_j'(x_j)|^s}{1 - T_j'(x_j)}.$$

Thus

$$\operatorname{tr}(\mathcal{L}_s) = \sum_j \frac{|T_j'(x)|^s}{1 - T_j'(x_j)} = \sum_{Tx = x} \frac{|T'(x)|^{-s}}{1 - (T'(x))^{-1}}. \quad \blacksquare$$

5.1. Strategy for super-exponential bounds. We can associate to the operators $\mathcal{L}_s : B \to B$ (on bounded analytic functions with the supremum norm $\|\cdot\|$) a sequence of real numbers defined as follows.

DEFINITION 5.2. We define the *approximation numbers* by

$$s_n(\mathcal{L}_s) = \inf\{\|\mathcal{L}_s - K\| : K = \text{operator with } n\text{-dimensional range}\}$$

for $n \geq 1$, where the infimum is taken over all linear operators $K : B \to B$ whose range is a finite dimensional space.

This definition makes sense for any bounded linear operator. However, the approximation numbers are crucial to getting bounds on the zeta functions [8].

5.1.1. *Bounds on the approximation numbers.* We can now explain the ideas behind the first ingredient. Let us replace $B(U)$, the space of bounded analytic functions, by $\mathcal{A} = \mathcal{A}(U)$, the space of analytic functions on U which are square integrable. We then write

$$\langle f, g \rangle = \int_U f g d(vol).$$

LEMMA 5.3. *We can bound* $s_n(\mathcal{L}_s) \leq C(s)\theta^{n+1}$ *where*

$$C(s) = \frac{\|\mathcal{L}_s\|_{\mathcal{A}(U) \to \mathcal{A}(U^+)}}{1 - \theta}$$

where:

1. U^+ *is a disk centred at* 0 *of radius* r*; and*
2. U *is a disk centred at* 0 *of radius* θr*.*

Proof. For $w \in \mathcal{A}(U)$ we write

$$\mathcal{L}_s w(z) = \sum_{k=0}^{\infty} l_k(w) z^k \in \mathcal{A}(U^+).$$

Since $\{z^k\}_{k=0}^{\infty}$ are orthogonal on $\mathcal{A}(U^+)$,

$$\langle \mathcal{L}_s w, z^k \rangle_{\mathcal{A}(U^+)} = l_k(w) \|z^k\|_{\mathcal{A}(U^+)}.$$

Thus by Cauchy-Schwarz,

$$|l_k(w)| \leq \|L_s w\|_{\mathcal{A}(U^+)} / \|z^k\|_{\mathcal{A}(U^+)}. \tag{5.1}$$

We can define a finite rank approximation by

$$\mathcal{L}_s^{(n)} w(z) = \sum_{k=0}^{n} l_k(w) z^k \in \mathcal{A}(U^+), \qquad n \geq 1.$$

Then

$$\|\mathcal{L}_s - \mathcal{L}_s^{(n)}\|_{\mathcal{A}(U)} \leq \sum_{k=n+1}^{n} |l_k(w)| \cdot \|z^k\|_{\mathcal{A}(U)} \leq \sum_{k=n+1}^{\infty} \|\mathcal{L}_s\|_{\mathcal{A}(U^+)} \frac{\|z^k\|_{\mathcal{A}(U)}}{\|z^k\|_{\mathcal{A}(U^+)}}$$

using (5.1). But we can compute

$$\|z^k\|_{\mathcal{A}(U^+)} = \sqrt{\frac{\pi}{k+1}} r^k \quad \text{and} \quad \|z^k\|_{\mathcal{A}(U)} = \sqrt{\frac{\pi}{k+1}} \theta^k r^k$$

and $\|\mathcal{L}_s\|_{\mathcal{A}(U^+)} \leq \|\mathcal{L}_s\|_{\mathcal{A}(U)\to\mathcal{A}(U^+)}.\|w\|_{\mathcal{A}(U)}$. Thus, since the definition of $s_n(\mathcal{L}_s)$ involves the infimum over all finite range operators K (including $\mathcal{L}_s^{(n)}$), we deduce that

$$s_n(\mathcal{L}_s) \leq \frac{\|\mathcal{L}_s\|_{\mathcal{A}(U)\to\mathcal{A}(U^+)}}{1-\theta}\theta^{n+1}.$$

This completes the proof. ∎

5.1.2. *Euler bounds.* We next give some simple but useful inequalities [26]. The first gives a simple but effective estimate on the terms in the tail of the series.

LEMMA 5.4. *Assume $s_n \leq C\theta^n$. For c_m defined by*

$$\prod_{n=0}^{\infty}(1+zs_n) = 1 + \sum_{m=1}^{\infty} c_m z^m, \quad z \in \mathbb{C},$$

we can bound $|c_m| \leq BC^m\theta^{m(m+1)/2}$ where $B = \prod_{n=1}^{\infty}(1-\theta^n) < +\infty$.

Proof. Since $c_m = \sum_{i_1<\cdots<i_m} s_{i_1}\cdots s_{i_m}$, for $m \geq 1$, we can bound

$$|c_m| \leq C^m \sum_{i_1<\cdots<i_m} \theta^{i_1+\cdots+i_m}.$$

We can prove by direct evaluation that

$$\sum_{i_1<\cdots<i_m} \theta^{i_1+\cdots+i_m} = \frac{\theta^{m(m+1)/2}}{(1-\theta)(1-\theta^2)\cdots(1-\theta^m)}. \blacksquare$$

We can also consider a bound on the coefficients in the power series for $\det(I - z\mathcal{L}_s)$. By Cauchy's theorem, if

$$\det(I - z\mathcal{L}_s) = 1 + \sum_{n=1}^{\infty} b_n z^n$$

then for $|z| = r$,

$$|b_n| = \left|\frac{1}{2\pi i}\int_{|\xi|=r} \frac{\det(I-\xi\mathcal{L}_s)}{\xi^{n+1}}d\xi\right| \leq \frac{1}{r^n}\sup_{|\xi|=r}|\det(I-\xi\mathcal{L}_s)|.$$

5.1.3. *Bounds on the coefficients.* The next bound relating the approximation numbers $\{s_n\}$ to the eigenvalues $\{\lambda_n\}$ is a classical result originally proved by Weyl for any compact operator defined on a Hilbert space.

LEMMA 5.5 (Weyl's Inequality). *If $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \cdots$ then*

$$\left|\prod_{j=1}^{n}\lambda_j\right| \leq \prod_{j=1}^{n} s_j.$$

We also need the following standard inequality.

LEMMA 5.6 (Hardy-Littlewood-Pólya). *Let $\{a_n\}$, $\{b_n\}$ be nonincreasing sequences of real numbers such that:*

1. $\sum_{j=1}^{n} a_j \leq \sum_{j=1}^{n} b_j$ *for $n \geq 1$; and*
2. $\Phi : \mathbb{R} \to \mathbb{R}$ *is convex.*

Then $\sum_{j=1}^{n}\Phi(a_j) \leq \sum_{j=1}^{n}\Phi(b_j)$.

We can make use of the Hardy-Littlewood-Pólya lemma as follows. Let

$$a_j = \log|\lambda_j|, \quad b_j = \log|s_j|, \quad \Phi(x) = \log(1+rx).$$

If $|z| = r$ then

$$\begin{aligned}
|\det(I - z\mathcal{L}_s)| &\le \prod_{j=1}^{\infty}(1+|z|\lambda_j) \\
&\le \prod_{j=1}^{\infty}(1+|z|s_j) \qquad \text{(by Lemmas 5.5 and 5.6)} \\
&\le 1 + B\sum_{m=1}^{\infty}(|z|C)^m \theta^{m(m+1)/2} \qquad \text{(by Lemma 5.4).}
\end{aligned}$$

Let $r = r(n) = \theta^{-n/2}/C$. Then

$$(Cr)^m \theta^{m^2/2} \le \begin{cases} \theta^{n^2/2} & \text{for } 1 \le m \le [n/2], \\ (\theta^{n/2})^m & \text{for } m > [n/2]. \end{cases}$$

Thus we can bound

$$|b_n| \le [n/2]\theta^{n^2/2} + \frac{\theta^{n^2/4}}{1-\theta^{n/2}} = O(\Theta^{n^2/2}) \quad \text{for any } \theta < \Theta < 1.$$

6. Anosov flows and geodesic flows. We can apply the previous ideas on zeta functions to the particular case of properties of Anosov flows. This includes the important classical case of geodesic flows on negatively curved surfaces. The main distinction is that we prefer to work in the setting of C^∞ systems rather than C^ω. This requires modifying the space of functions upon which the transfer operates (and ultimately changing the operator itself).

In particular, we can consider for Anosov flows two types of problems: rates of mixing and error terms in counting closed orbits. We begin with the definition.

Let $\phi_t : M \to M$ be C^∞ flow on compact manifold.

DEFINITION 6.1. We call $\phi_t : M \to M$ *Anosov* if there exists a $D\phi$-invariant splitting $TM = E^0 \oplus E^s \oplus E^u$ such that:

1. E^0 is a one dimensional bundle tangent to the flow; and
2. there exist $C, \lambda > 0$ such that
$$\|D\phi_t|E^s\| \le Ce^{-\lambda t} \quad \text{and} \quad \|D\phi_{-t}|E^s\| \le Ce^{-\lambda t}$$
for $t \ge 0$ [2].

We recall the classical example of an Anosov flow on a three dimensional manifold provided by geodesic flows on surfaces.

EXAMPLE 6.2 (Classic example). Let $M = SV$ be the three dimensional unit tangent bundle for a compact surface V of curvature $\kappa < 0$. Given $v \in M$ we can consider the unique unit speed geodesic $\gamma_v : \mathbb{R} \to V$ with $\dot\gamma_v(0) = v$. We then define the geodesic flow $\phi_v : M \to M$ by $\phi_t(v) = \dot\gamma_v(t)$.

Let us henceforth concentrate on the particular case of geodesic flows, for which we can prove stronger results. We shall consider the rate of mixing, and in a later section describe the closely related asymptotic estimates on the number of closed orbits (or equivalently closed geodesics).

Let m be the Liouville (or SRB) measure for ϕ. This is the unique invariant measure equivalent to the volume on $SV = M$. As is well known the geodesic flow is ergodic with respect to m. However, it is also known that the flow is (strong) mixing with respect to m. We recall a useful definition.

Definition 6.3. Let $F, G : M \to \mathbb{R}$ be C^∞ and define the correlation function by

$$\rho(t) := \int F \circ \phi_t G dm - \int F dm - \int G dm$$

for $t \geq 0$.

The flow is strong mixing because $\rho(t) \to 0$ for any C^∞ functions F, G (or equivalently, for $F, G \in L^2(m)$).

However, a much stronger result is known on the speed of convergence to zero of $\rho(t)$. This is presented as the following theorem, which deals with the first of two intimately related properties [21].

Theorem 6.4 (Dolgopyat: Exponential mixing). *Let $\phi_t : M \to M$ be the geodesic flow on a compact surface of (variable) negative curvature. There exists $\epsilon > 0$ such that for all $F, G \in C^\infty(M)$ there exists $C > 0$ with*

$$|\rho(t)| \leq C e^{-\epsilon t} \quad \textit{for } t \geq 0.$$

This famous result is due to D. Dolgopyat and is now 20 years old, but because of the technical nature of the proof it still remains a little mysterious to many people. A more geometric formulation, which works better for geodesic flows on higher dimensional manifolds, was given by C. Liverani [40].

We shall briefly describe the original proof, which uses Markov sections and transfer operators in a C^1 setting. Although this particular approach is perhaps a little old fashioned, it fits in well with our preceding analysis of iterated function schemes. We will also concentrate on the three dimensional case for simplicity. The choice of Markov sections for the flow is then done by analogy with the well known approach of Adler-Weiss constructing Markov partitions for linear hyperbolic toral automorphisms [1], [56]. There one uses the stable and unstable manifolds for a fixed point to give the boundaries of the Markov partition and for geodesic flows one uses the weak stable and unstable manifolds associated to a closed orbit for the flow.

Step 1. Let $\dim M = 3$ and let τ be a closed orbit for ϕ. We can define the weak stable and unstable manifolds for τ, which are two dimensional immersed submanifolds

$$W^s(\tau) = \{x \in M : d(\phi_t x, \tau) \to 0 \text{ as } t \to +\infty\},$$
$$W^u(\tau) = \{x \in M : d(\phi_{-t} x, \tau) \to 0 \text{ as } t \to +\infty\}.$$

(These are weak stable and unstable manifolds for the closed orbit τ.) In practice we will only want to consider parts of $W^s(\tau)$ and $W^u(\tau)$ which are a bounded distance (along the submanifolds) to the original orbit τ. We also introduce sections S_i transverse to the flow (with boundaries contained in $W^s(\tau)$ and $W^u(\tau)$) which help divide M into flow boxes P_i, say, for $i = 1, \ldots, m$. We can view these as parallelepipeds of the form

$$P_i = \{\phi_t w : 0 \leq t \leq r_i(w)\}, \quad i = 1, \ldots, n,$$

where $r_i : S_i \to \mathbb{R}^+$.

STEP 2. We can now define a discrete map. This is first achieved by identifying the flow boxes along the the leaves of a suitable foliation. More precisely, we can define the one dimensional stable manifolds:

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

for each $x \in M$. The following classical result helps explain why we can work in the C^1 setting.

LEMMA 6.5 (Hopf, Hirsch-Pugh). *For geodesic flows on surfaces the family* $\{W^{ss}(x)\}_{x \in M}$ *gives a* C^1 *foliation of* M [32].

STEP 3. We can now introduce an associated C^1 one dimensional expanding map. "Identifying" sections S_i along stable manifolds gives a one dimensional C^1 manifold or "interval".

We begin with the natural projection $P_i \to S_i$ from each three dimensional parallelepiped to the corresponding two dimensional section along the orbits of the flow. We also have the following useful trick to relate the C^1 nature of the foliations to the sections [58].

LEMMA 6.6 (after Ruelle). *We choose the sections* S_i *so that they (and thus the parallopipeds* P_i*) are foliated by strong stable manifolds.*

The Poincaré map between sections gives a C^1 map $T : \bigcup_i I_i \to \bigcup_i I_i$. The return (or transition) time between sections gives a C^1 function $r : \bigcup_i I_i \to \mathbb{R}^+$.

STEP 4. We can construct invariant measures (following Bowen-Ruelle). Let $\psi : I \to \mathbb{R}$ be a Hölder continuous function (used as a potential to define a Gibbs measure).

DEFINITION 6.7. We can define the Gibbs measure (or equilibrium state) μ_ψ:

$$h(\mu_\psi) + \int \psi d\mu_\psi = \sup\left\{ h(\mu) + \int \psi d\mu : \mu = T\text{-invariant} \right\} =: P(\psi)$$

where $P(\psi)$ is the pressure function for ψ.

The measures μ_ψ on $\bigcup_i I_i$ correspond to a flow invariant measure m on M, which is given by a simple construction [14]:

1. we can extend μ_ψ on I to $\bar{\mu}_\psi$ on $\bigcup_i S_i$ (the natural extension);
2. we can extend $\bar{\mu}_\psi$ to a ϕ-invariant measure m on M by

$$dm = \frac{d\bar{\mu}_\psi \times dt}{\int r d\bar{\mu}_\psi}$$

where $m(\partial P_i) = 0$.

Of course, we can consider particular choices of Hölder continuous potentials. These give rise to different invariant measures for the geodesic flow.

EXAMPLE 6.8. Let $\psi : I \to \mathbb{R}$. Then

1. if $\psi(x) = -\log|T'(x)|$ then m is the Liouville measure; and
2. if $\psi(x) = -hr$, where h is the topological entropy of the flow, then m is the measure of maximal entropy (or Bowen–Margulis measure).

STEP 5. We can now introduce transfer operators. Let $C^1(I)$ be the Banach space of C^1 functions $w : I \to \mathbb{C}$ with norm $\|w\| = \|w\|_\infty + \|w'\|_\infty$.

We can understand the properties of the measures μ_ϕ (and thus of the corresponding measure $\overline{\mu}_\phi$ and flow invariant measure m) through the spectral properties of an associated transfer operator.

DEFINITION 6.9. Let $\psi : I \to \mathbb{R}$ be a C^1 function. Then we can define the transfer operator $\mathcal{L}_\psi : C^1(I) \to C^1(I)$ by

$$\mathcal{L}_\psi w(x) = \sum_{y\colon Ty=x} e^{\psi(y)} w(y).$$

We can now describe the properties of this operator [59], [13], [47].

THEOREM 6.10 (Ruelle). *Let $\psi : I \to \mathbb{R}$ be C^1.*

1. *$\mathcal{L}_\psi$ has a (maximal) positive eigenvalue $e^{P(\psi)}$ (and a positive eigenvector h_ψ).*
2. *The dual operator $\mathcal{L}_\psi^* : C^1(I)^* \to C^1(I)^*$ (defined by $\mathcal{L}_\psi^*\nu(w) = \nu(\mathcal{L}_\psi w)$ for $\nu \in C^1(I)^*$ and $w \in C^1(I)$) has an eigenmeasure ν_ψ, i.e., $\mathcal{L}_\psi^*\nu_\psi = e^{P(\psi)}\nu_\psi$.*
3. *If $\sup_{x\in I} 1/|T'(x)| < \theta < 1$, say, then $\mathcal{L}_\psi : C^1(I) \to C^1(I)$ has only isolated eigenvalues outside the disk of radius $\theta e^{P(\psi)}$.*

Recall that in the previous context of C^ω functions the operator $\mathcal{L}_\psi$ was nuclear, and thus had countably many eigenvalues. But since we now have to work in the C^1 category, there may be more eigenvalues, although part 3 of the above result implies that they don't occur outside of the disk of radius $\theta e^{P(\psi)}$.

Part 1 of the theorem allows us to make a particularly useful simplification [47].

COROLLARY 6.11 (Normalization). *Given $\psi \in C^1(I)$ we define $\overline{\psi} = \psi + \log h_\psi - \log h_\psi \circ T - P(\psi)$. Then*

1. *$\mathcal{L}_{\overline{\psi}}1 = 1$, the constant function with value 1;*
2. *$\mathcal{L}_{\overline{\psi}}^*\nu_{\overline{\psi}} = \nu_{\overline{\psi}}$, and then $\nu_{\overline{\psi}} = \mu_\psi$, the Gibbs measure for ψ*

STEP 6. Finally, we have a strategy for proving "statistical properties", such as exponential mixing, for the original flow. Let μ be a ϕ-invariant Gibbs measures and $F, G \in C^\infty(M)$. We have the Laplace transform

$$\widehat{\rho}(s) = \int_0^\infty e^{-st}\rho(t)dt, \qquad s \in \mathbb{C},$$

which converges for $Re(s) > 0$. We want to apply the following result to convert properties of $\widehat{\rho}(s)$ into bounds on $\rho(t)$ [57].

Theorem 6.12 (Paley-Wiener). *Assume we can show $\widehat{\rho}(s)$ has an analytic extension to $Re(s) \geq -\epsilon_0$ say, and*

$$\sup_{-\epsilon_0 \leq \delta \leq 0} \left| \int_0^\infty \widehat{\rho}(\delta + it) dt \right| < +\infty$$

for some $\epsilon_0 > 0$. Then for any $0 < \epsilon < \epsilon_0$ there exists $C > 0$ such that $|\rho(t)| \leq Ce^{-\epsilon t}$ for $t \geq 0$.

What remains is to modify the transfer to include the complex variable $s \in \mathbb{C}$ and to write $\widehat{\rho}(s)$ in terms of this. We will discuss this in the next section.

7. The complex transfer operator. Given C^1 functions $\psi, r : I \to \mathbb{R}$ and $s \in \mathbb{C}$ we can define a complex transfer operator $\mathcal{L}_{\psi - sr} : C^1(I) \to C^1(I)$ by

$$\mathcal{L}_{\psi - sr} w(x) = \sum_{Ty = x} e^{(\psi - sr)(y)} w(y).$$

Remark 7.1. When $s = 0$, this reduces to the usual "real" operator.

Usually it is convenient to assume $\mathcal{L}_{\psi - hr} 1 = 1$ where 1 denotes the constant function 1 (and then $\mathcal{L}^*_{\psi - hr} \mu_{\psi - \sigma r} = \mu_{\psi - hr}$ is a Gibbs measure for $\psi - hr$). In fact, we can usually assume this without loss of generality, by Corollary 6.11.

The following is a partial analogue of Theorem 6.10 for the operator $\mathcal{L}_{\psi - sr}$ [49], [47].

Theorem 7.2 (Complex Ruelle Operator Theorem). *Let $s = \sigma + it$. Then*

1. *The spectral radius of $\mathcal{L}_{\psi - \sigma r}$ satisfies $\rho(\mathcal{L}_{\psi - sr}) \leq e^{P(\psi - \sigma r)}$.*
2. *$\mathcal{L}_{\psi - sr} : C^1(I) \to C^1(I)$ has only isolated eigenvalues outside $\theta e^{P(\psi - \sigma r)}$.*

We can now try to relate the transfer operator $\mathcal{L}_{\psi - \sigma r}$ to the Laplace transform $\widehat{\rho}(s)$. The spectral properties of the operator then lead to properties of the complex function.

Claim 7.3. *We have the following properties.*

1. *There exists $\epsilon > 0$ such that $\widehat{\rho}(s)$ has a meromorphic extension to $Re(s) > -\epsilon$.*
2. *If $s = s_0$ is a pole for $\widehat{\rho}(s)$ then 1 is an eigenvalue for $\mathcal{L}_{\psi - \sigma_0 r}$.*

We briefly recall the idea of the proof of the claim. We want to write

$$\widehat{\rho}(s) = \int_I f_s \Big(\sum_{n=0}^\infty \mathcal{L}^n_{\psi - \sigma r} g_{-s} \Big) d\mu(x)$$

where $\sum_{n=0}^\infty \mathcal{L}^n_{\psi - \sigma r} = (1 - \mathcal{L}_{\psi - \sigma r})^{-1}$ for suitable functions f_s, g_{-s}. If we can replace the functions F and G by functions which are constant on stable leaves in the parallelepiped then we could associate

$$F \mapsto f_s(x) = \int_0^{r(x)} e^{-st} F(x,t) dt \in C^\alpha(I),$$

$$G \mapsto g_{-s}(x) = \int_0^{r(x)} e^{st} F(x,t) dt \in C^\alpha(I).$$

The justification for this comes from a result of Ruelle.

All of the above framework was in place in the 1980s. However, it took another decade for this to be used to deduce exponential decay of correlations.

To apply the Paley-Wiener theorem we need control on the eigenvalues of $L_{\psi-\sigma r}$ (i.e., poles of $\widehat{\rho}(s)$). This is achieved by the following famous result of Dolgopyat [21].

THEOREM 7.4 (Dolgopyat). *There exist $\epsilon > 0$ and $0 < \rho < 1$ so that for $s = \sigma + it$:*

1. *$\mathcal{L}_{\psi-sr} : C^1(I) \to C^1(I)$ (or $\mathcal{L}_{\psi-sr} : C^\alpha(I) \to C^\alpha(I)$) has spectral radius $\rho(L_{\psi-sr}) \le \rho$ whenever $\sigma > -\epsilon$ and $t > \epsilon$; and*
2. *there exist $C > 0$ and $A > 0$ so that whenever $\sigma > -\epsilon$, $|t| > \epsilon$ and*

$$n = k[A \log |t|] + l \quad \text{for } k \ge 0 \text{ and } 0 \le l \le [A \log |t|] - 1$$

then $\|\mathcal{L}^n_{\psi-sr}\| \le C\rho^{k[A \log |t|]}$.

Having outlined the way in which properties of the transfer operator lead to the dynamical properties of the geodesic flow, the following question remains.

QUESTION. What properties does the geodesic flow have which are needed for the result? How do they filter through to the transfer operator?

The geometric features of geodesic flow can be encoded into the Markov sections and their collapsed versions.

8. Uniform bounds on transfer operators. In this section we outline the key ideas in the proof of Dolgopyat's estimate.

8.1. A sketch of the proof. We want to define a C^1 function $\Delta : I \to \mathbb{R}$ of the form $\Delta(x) = r(y) - r(z)$ where y, z are preimages of x under the expanding map, i.e., $Ty = Tz = x$.

We then have a function defined locally (in a neighbourhood of x_0 with distinct preimages y_0, z_0, i.e., $T(y_0) = T(z_0) = x_0$) by

$$\Delta(x) = (r(y) - r(z)) - (r(y_0) - r(z_0)).$$

We can assume that $I \ni x \mapsto \Delta(x)$ is C^1 and there exists $C > 0$ such that locally we can write

$$\frac{1}{C} \le \frac{\Delta(x)}{x - x_0} \le C.$$

This is essentially all that is required from the flow.[2]

Sketch proof of Dolgopyat's theorem. We want to show that $\mathcal{L}_{\psi-sr}$ is a C^1-contraction. Actually, this is achieved by a series of steps:

(i) showing that $\mathcal{L}_{\psi-sr}$ is a L^1-contraction;
(ii) showing that $\mathcal{L}_{\psi-sr}$ is a L^1-contraction implies it is a C^0-contraction; and
(iii) showing that $\mathcal{L}_{\psi-sr}$ is a C^0-contraction implies it is a C^1-contraction (or C^α-contraction).

[2] In practice, we need to take higher iterates of T.

This is a form of "bootstrapping argument" whereby we improve the regularity step by step.

We will consider each of these steps (in reverse order) where $w \in C^1(I)$:

Sketch of part (iii). Assume we already have a C^0 estimate: There exists $0 < \theta_0 < 1$ such that

$$\|\mathcal{L}^n w\|_\infty = O(\theta_0^n). \tag{8.1}$$

Then we can use the following important bound.

LEMMA 8.1 (after Doeblin-Fortet, Lasota-Yorke). *There exist $C>0$ and $\|1/T'\|_\infty<\theta<1$ such that*

$$\|\mathcal{L}^n_{\psi-sr} w\| \leq C|t|\|w\|_\infty + \theta^n \|w\| \tag{8.2}$$

for all $n \geq 1$, where $s = \sigma + it$.

Applying (8.2) twice we can write

$$\|\mathcal{L}^{2n}_{\psi-sr} w\| = \|\mathcal{L}^n_{\psi-sr}(\mathcal{L}^n_{\psi-sr} w)\| \leq C|t|\|\mathcal{L}^n_{\psi-sr} w\|_\infty + \theta^n (C|t|\|w\|_\infty + \theta^n \|w\|)$$

where $\|\mathcal{L}^n_{\psi-sr} w\|_\infty = O(\theta_0^n)$ by (8.1) and $C|t|\|w\|_\infty + \theta^n\|w\|$ is uniformly bounded. Thus

$$\|\mathcal{L}^{2n}_{\psi-sr} w\| = O(|t|\theta_1^n)$$

where $\theta_1 = \max(\theta, \theta_0)$.

Sketch of part (ii). Assume we had L^1-estimates

$$\|\mathcal{L}^n_{\psi-sr} w\|_{L^1} = \int |\mathcal{L}^n_{\psi-sr} w| \mu_\sigma = O(\theta_2^n) \tag{8.3}$$

for some $0 < \theta_2 < 1$, where $\mathcal{L}_{\psi-\sigma r}\mu_\sigma = \mu_\sigma$ is the Gibbs measure for $\psi - \sigma r$.

By Theorem 6.10 (i.e., the existence of a spectral gap for $\mathcal{L}_{\psi-\sigma r}$) there exists $0 < \theta_3 < 1$ such that

$$\left\| \mathcal{L}_{\psi-\sigma r} w - \int w d\mu_\sigma \right\|_\infty = O(\theta_3^n).$$

Thus for $n \geq 1$:

$$\|\mathcal{L}^{2n}_{\psi-\sigma r} w\|_\infty = \|\mathcal{L}^n_{\psi-\sigma r}(\mathcal{L}^n_{\psi-\sigma r} w)\|_\infty \leq \int |\mathcal{L}^n_{\psi-\sigma r} w| d\mu_\sigma + O(\theta_3^n),$$

and using (8.3) we get that $\|\mathcal{L}^{2n}_{\psi-\sigma r} w\|_\infty = O(\theta_4^n)$ where $\theta_4 := \max\{\theta_2, \theta_3\}$.

Finally, "all" that remains is an argument to get L^1-contraction (somehow using the properties of $\Delta(x)$).

Sketch of part (i). The basic idea is that the operator contracts in the L^1 norm because of cancellations that arise because of differences in the arguments that can occur in the various terms arising from $\mathcal{L}_{\psi-s\phi}$. The important thing is that this should be uniform in $t = |Im(s)|$ to ensure that the Laplace transform has an analytic extension to a uniform strip.

More precisely, we can summarize the idea as follows:

(a) $\mathcal{L}_{\psi-\sigma r} w(x)$ contains contributions from two terms

$$e^{\psi(y)-\sigma r(y)} e^{-itr(y)} + e^{\psi(z)-\sigma r(z)} e^{-itr(z)}$$

with $Ty = Tz = x$ and where the difference in the arguments of the two terms is obviously $t(r(y) - r(z)) = t\Delta(x) \pmod{2\pi}$.

(b) In particular, when $\frac{\pi}{2} \leq t\Delta(x) \leq \frac{3\pi}{2} \pmod{2\pi}$ a little trigonometry shows that

$$|\mathcal{L}_{\psi - sr} w(x)| \leq \beta |\mathcal{L}_{\psi - \sigma r} w(x)|$$

for some $0 < \beta < 1$ (which is independent of t).

(c) For each sufficiently large t we can divide I into a union of (small) subintervals $\{I_i\}$ of length $|I_i| \asymp 1/|t|$ consisting of:

(i) *Good intervals.* These are intervals I_i for which $x \in I_i$ implies that $t\Delta(x) \in [\pi/2, 3\pi/2]$. Thus by (b) above, if I_i is a good interval and $x \in I_i$ then

$$|\mathcal{L}_{\psi - sr} w(x)| \leq \beta |\mathcal{L}_{\psi - \sigma r} w(x)|.$$

(ii) *Bad intervals.* These are simply the complements of the good intervals and here we just use the trivial inequality

$$|\mathcal{L}_{\psi - sr} w(x)| \leq |\mathcal{L}_{\psi - \sigma r} w(x)|.$$

A natural question to ask at this stage is: *What do we use about μ and what properties does it have which lead to a uniform contraction?* We will now address this.

(d) Although as t increases one expects more good (and bad) intervals, the total measure of their union is (uniformly) bounded away from zero. In particular, the uniform contractions on the good intervals then lead to a uniform contraction in the L^1-norm.

To see this crucial feature, we can compare the measures of each good interval I_i and one of its neighbouring bad intervals I_{i+1}, say. The important thing about the measure is that it has the "doubling property": there exist $A, B > 0$ such that providing $|t|$ is sufficiently large we can bound $A \leq \mu(I_i)/\mu(I_{i+1}) \leq B$ for all such intervals I_i and I_{i+1}.

We can therefore conclude that providing t is sufficiently large we can bound

$$|\mathcal{L}_{\psi - sr} w(x)| \leq \beta |\mathcal{L}_{\psi - \sigma r} w(x)|$$

on a set of uniformly bounded (from below) measure. This implies contraction in L^1-norm.

This completes our sketch of the basic argument of Dolgopyat. However, at the risk of obscuring the basic idea with too much detail, let us flesh out part (d) a little more.

8.2. More details on the proof. *A more elaborate account of part (d).* For notational convenience we denote

$$\|h\| = \max\{\|h\|_\infty, \|h'\|/|t|\}$$

and consider two cases: one very easy, and the other less so.

(I) *Easy case.* Assume $2C|t| \cdot |h|_\infty \leq |h'|_\infty$ where C is the constant from Lemma 8.1. We can fix $\frac{1}{2} < \eta < 1$ and then choose k such that $\frac{1}{2} + \theta^k < \eta$. Then by Lemma 8.1 we have

$$\frac{1}{|t|} |(\mathcal{L}^k h)'|_\infty \leq C|h|_\infty + \frac{\theta^k}{|t|} \leq (1/2 + \theta^k) \leq \eta \|h\|$$

by hypothesis and definition of $\|\cdot\|$, i.e., $\|\cdot\|$ contracts (in this case).

This still leaves the other case.

(II) *Difficult case.* Assume $2C|t||h|_\infty \geq |h'|_\infty$. We want to choose a sequence of C^1 functions $u_n : I \to \mathbb{R}$, $n \geq 0$, such that the following properties hold:

1. $0 \leq |v_n| \leq u_n$ for $v_n := \mathcal{L}^n_{\psi - sr} h$, $n \geq 1$;
2. there exists $0 < \beta < 1$ with $\|u_n\|_2 \leq \beta^n$, $n \geq 1$;
3. $|u'_n/u_n| \leq 2C|t|$, $n \geq 1$; and
4. $|v'_n/u_n| \leq 2C|t|$, $n \geq 1$.

The functions u_n have the advantage over v_n of being real valued. The existence of such functions u_n comes from an iterative construction. Let $u_0 = 1$, say. Assume u_n has been constructed. We need a "calculus lemma" relating u_n to v_n.

Lemma 8.2 (Calculus Lemma). *There exist $0 < \eta < 1$, $\epsilon > 0$, $\delta > 0$ such that for all $x_0 \in I$ there exists a nearby interval $[x_1 - \delta/|t|, x_1 + \delta/|t|]$ with $|x_1 - x_0| \leq \epsilon/|t|$ such that for all x in this interval we have either*

$$|e^{-sr(y)} v_n(y) + e^{-sr(z)} v_n(z)| \leq \eta e^{-\sigma r(y)} u_n(y) +^{-\sigma r(z)} u_n(z)$$

or

$$|e^{-sr(y)} v_n(y) + e^{-sr(z)} v_n(z)| \leq \eta e^{-\sigma r(z)} u_n(z) +^{-\sigma r(y)} u_n(y).$$

We can choose (reasonably good) intervals

$$[x_0, x_1], [x_2, x_3], \ldots, [x_{2n-2}, x_{2n-1}]$$

upon which one of the two inequalities in Lemma 8.2 hold. We then continue to define the sequence of functions iteratively by

$$u_{n+1}(x) = \mathcal{L}_{\psi - \delta r}(u_n \chi)(x)$$

where

$$\chi(x) = \begin{cases} \eta & \text{if } x_{2n} - \frac{x_{2n+1} - x_{2n}}{4} < x < x_{2n} + \frac{x_{2n+1} - x_{2n}}{4}, \\ 1 & \text{if } x_{2n+1} \leq x \leq x_{2n+2}, \\ & \text{a smooth interpolation in-between,} \end{cases}$$

with $|\chi|_\infty \leq 1$ and $|\chi'|_\infty \leq E|t|\chi(x)$. By construction we then have

$$|u'_{n+1}| = |(\mathcal{L}_{\psi - \sigma r}(u_n \chi))'| \leq C|t||(u_n \chi)| + \theta|(u_n \chi)')|$$

and by the chain rule

$$|(u_n \chi)')(x)| \leq |u'_n(x)|\chi(x) + u_n(x).|\chi'(x)| \leq (2C|t|u_n(x))\chi(x) + u_n(x)(E|t|\chi(x)).$$

Combining these bounds we have $|u'_{n+1}(x)| \leq 2C|t||u_{n+1}(x)|$ (providing $0 < \theta < 1$ is sufficiently small) i.e., 3. holds for u_{n+1}. Moreover,

$$\begin{aligned} |v'_{n+1}(x)| = |(\mathcal{L}_{\psi - \sigma r} v_n)'(x)| &\leq C|t|\mathcal{L}_{\psi - \sigma r}|v_n(x)| + \theta \mathcal{L}_{\psi - \sigma r}|v'_n(x)| \\ &\leq C|t|\mathcal{L}_{\psi - \sigma r} u_n(x) + \theta \mathcal{L}_{\psi - \sigma r} u'_n(x) \leq 2C|t|u_{n+1}(x), \end{aligned}$$

i.e., 4. holds for u_{n+1}.

To establish 2. it suffices to show that there exists $0 < \beta < 1$ such that $\|u_{n+1}\|_2 \leq \beta\|u_n\|_2$ for all $n \geq 0$. Moreover, this is (essentially) what we need to complete the proof of the theorem since then

$$\|\mathcal{L}^n_{\psi - \sigma r} h\|_2 \leq \|u_n\|_2 \leq \beta^n$$

for $n \geq 1$. To this end, observe that if $x \in [x_{2i+1}, x_{2i+2}]$ then

$$u_{n+1}^2(x) = (\mathcal{L}_{\psi-\sigma r}(\chi u_n)(x))^2 \leq (\mathcal{L}_{\psi-\sigma r}(\chi^2)(x))(\mathcal{L}_{\psi-\sigma r}(u_n^2)(x))$$

where

$$\mathcal{L}_{\psi-\sigma r}(\chi^2)(x) \leq \beta_0 < 1.$$

Thus (on these good intervals)

$$\int_{x_{2i+1}}^{x_{2i+2}} u_{n+1}^2(x) d\nu(x) \leq \beta_0 \int_{x_{2i+1}}^{x_{2i+2}} \mathcal{L}_{\psi-\delta r} u_n^2(x) d\nu(x)$$

and we can trivially bound (on the bad intervals)

$$\int_{x_{2i}}^{x_{2i+1}} u_{n+1}^2(x) d\nu(x) \leq \int_{x_{2i}}^{x_{2i+1}} \mathcal{L}_{\psi-\delta r} u_n^2(x) d\nu(x).$$

But for $x' \in [x_{2i}, x_{2i+1}]$ and $x'' \in [x_{2i+1}, x_{2i+2}]$ we have

$$\frac{u_{n+1}(x')^2}{u_{n+1}(x'')^2} \leq \exp\left(2 \int_{x'}^{x''} |(\log u_{n+1})'(x)| dx\right) \leq \exp(2|x_{2i+2} - x_{2i}|.2C|t|) \leq B$$

say.

Moreover,

$$\frac{\int_{x_{2i}}^{x_{2i+1}} u_n^2 d\nu}{\int_{x_{2i+1}}^{x_{2i+2}} u_n^2 d\nu} \leq B\left(\sup_i \left\{ \frac{\nu([x_{2i}, x_{2i+1}])}{\nu([x_{2i+1}, x_{2i+2}])} \right\}\right) \leq A$$

say. Thus

$$\int u_{n+1}^2 d\mu = \sum_i \beta_0 \int_{x_{2i}}^{x_{2i+1}} u_n^2 d\nu + \int_{x_{2i+1}}^{x_{2i+2}} u_n^2 d\nu \leq \beta^2 \int u_n^2 d\nu$$

for some $0 < \beta < 1$. □

Of course this method seems a little complicated and, perhaps, rather restricted in its application. This begs the question:

QUESTION. More generally, how useful are these ideas?

In fact, this basic method has been used in several different settings. For example:

(i) Baladi-Vallée used similar results on transfer operators to study statistical properties of (Euclidean) algorithms [7].

(ii) Avila-Gouëzel-Yoccoz showed exponential mixing for Teichmüller geodesic flows [5].

REMARK 8.3 (Teichmüller flows). Let V be a closed surface. Let $\mathcal{M}$ be the space of Riemann metrics g (moduli space). Let ρ be the Teichmüller metric on $\mathcal{M}$ with normalized volume $(vol)_\rho$.

Let $F, G : S\mathcal{M} \to \mathbb{R}$ be smooth (compactly supported) functions. Then

$$\rho(t) = \int F \phi_t G d(vol)_\rho - \int F d(vol)_\rho \int G d(vol)_\rho$$

tends to zero exponentially fast.

The method is based on modelling by a symbolic flow. A simpler example would be when $V = \mathbb{T}^2$; then the modular surface $\mathcal{M}$ is equal to $\mathbb{H}^2/PSL(2,\mathbb{Z})$ and the dynamics corresponds to (the natural extension of) the Gauss map $T : (0,1) \to (0,1)$ defined by

$T(x) = 1/x \pmod 1$ and a roof function $r : (0,1) \to \mathbb{R}$ defined by $r(x) = -2\log x$ and the volume $d(vol)_\rho = Cdxdt/(1+x)$.

REMARK 8.4 (Weil-Petersson flows). Another metric on $\mathcal{M}$ is the Weil-Petersson metric which has a nice dynamical interpretation (after McMullen [46]). To a family $g_\lambda \in \mathcal{M}$ of metrics $\lambda \in (-\epsilon, \epsilon)$ we can associate the geodesic flows $\phi_t^{g_\lambda} : S\mathcal{M} \to S\mathcal{M}$. Each can be modelled by a suspension of a subshift of finite type $\sigma : \Sigma \to \Sigma$ and a family of Hölder roof functions $r_\lambda : \Sigma \to \mathbb{R}$. If we write $r_\lambda = r_{\lambda_0} + (\lambda - \lambda_0)\dot{r}_{\lambda_0} + o(\lambda - \lambda_0)$ corresponding to the change in metric $g_\lambda = g_{\lambda_0} + (\lambda - \lambda_0)\dot{g}_{\lambda_0} + o(\lambda - \lambda_0)$ then we can write the Weil-Petersson metric or pressure metric where

$$\|\dot{g}_{\lambda_0}\|_{WP} = \frac{\partial^2}{\partial t^2} P(-r_0 + t\dot{r}_{\lambda_0})|_{t=0} > 0.$$

The ergodicity and mixing properties of the geodesic flow with this metric were studied by Burns-Masur-(Matheus)-Wilkinson [18], [17].

9. Counting closed geodesics. The same basic method leads to error terms in counting functions for the number of closed orbits (or equivalently closed geodesics) for the flow.

One can improve the famous Margulis estimate for lengths of closed geodesics γ:

$$\mathrm{Card}\{\gamma : l(\gamma) \leq T\} \sim e^{hT}/hT \quad \text{as } T \to +\infty$$

where h is the topological entropy of $\phi_{t=1}$ [41], [42].

The improvement is the exponential error term, once we get the correct principal term:

THEOREM 9.1 (Counting closed geodesics). *Let $\phi_t : M \to M$ be the geodesic flow on a compact surface of (variable) negative curvature. There exists $\epsilon > 0$ such that*

$$\mathrm{Card}\{\gamma : l(\gamma) \leq T\} = \int_2^{e^{hT}} \frac{1}{\log u} du + O(e^{(h-\epsilon)T}) \quad \text{as } T \to +\infty$$

where

$$\int_2^{e^{hT}} \frac{1}{\log u} du \sim e^{hT}/hT \quad \text{as } T \to +\infty.$$

This is a companion result to the exponential mixing for the geodesic flows. In place of the Laplace transform of the correlation function consider another complex function, the *Selberg zeta function*

$$Z(s) = \prod_{n=1}^{\infty} \prod_{\gamma} (1 - e^{-(s+n)l(\gamma)}), \quad s \in \mathbb{C}.$$

This converges for $Re(s) > h$. We can consider the logarithmic derivative

$$\frac{d}{ds} \log Z(s) = \frac{Z'(s)}{Z(s)} = \frac{-1}{s-h} + A(s)$$

where $A(s)$ is an analytic function for $Re(s) > h - \epsilon$, say.

Using Cauchy's theorem we can relate

$$\int_{Re(s)=\epsilon/2} \frac{Z'(s)}{Z(s)} ds$$

to $\pi(T) = \text{Card}\{\gamma : l(\gamma) \leq T\}$ and deduce Theorem 9.1 using a straightforward analysis borrowed from prime number theory [52].

We have formulated this in the context of compact surfaces V. However, the dynamical approach is much more flexible.

QUESTION. How can we generalize the Selberg zeta function?

Let us try to answer this question in the next two items.

(iii) *Thin groups.* Examples of "thin groups" are nonlattice subgroups of $PSL(2,\mathbb{R})$. Let us mention a recent result in this direction. Let $\Gamma < PSL(2,\mathbb{Z})$ be a subgroup. Let $\gamma_0 \in PSL(2,\mathbb{Z}/q\mathbb{Z})$ and let $\delta(\Gamma) = \delta$ be the Hausdorff dimension of the limit set. Bourgain-Gamburd-Sarnak [10] estimated

$$\text{Card}\{\gamma \in \Gamma : \|\gamma\| \leq T, \gamma = \gamma_0((\text{mod } q))\} = \frac{CT^{2\delta}}{\text{Card}PSL(2,\mathbb{Z}/q\mathbb{Z})} + \text{“error term”}$$

with an explicit error term. For $\frac{1}{2} < \delta \leq 1$ the proof uses the classical Laplacian. However, for $0 < \delta \leq \frac{1}{2}$ the proof uses transfer operator techniques.

(iv) *Higher Teichmüller theory.* Given a compact Riemann surfaces V with $\kappa = -1$ we recall that the surface V can be written as $\mathbb{H}^2/\Gamma$ where Γ are isometries of $\mathbb{H}^2$. A closed orbit (or closed geodesic) then corresponds to a conjugacy class $[g]$ in $\Gamma - \{e\}$. The length of the closed orbit γ is then given by $l(\gamma) = \cosh^{-1}(\text{tr}(g)/2)$. The Selberg zeta function for the Riemann surface V can be written as

$$Z_2(s) = \prod_{n=0}^{\infty} \prod_{\gamma} (1 - e^{-(s+n)l(\gamma)}),$$

where $s \in \mathbb{C}$. This has an analytic extension to $\mathbb{C}$. One natural generalization to Higher Teichmüller Theory and representations in $PSL(d,\mathbb{R})$ would involve $R([g]) \in PSL(d,\mathbb{R})$, where R is a representation in $PSL(d,\mathbb{R})$. In the case of an appropriate representation (in the so called Hitchin component) there exists a largest eigenvalue $e^{l(g)}$ of $R([g])$ [39] and we could again define the corresponding zeta function by

$$Z_d(s) = \prod_{n=0}^{\infty} \prod_{g} (1 - e^{-(s+n)l(g)})$$

where $s \in \mathbb{C}$. This too has a meromorphic extension to $\mathbb{C}$.

10. The newer approach to transfer operators. The traditional approach to transfer operators we have described in the previous sections has proved quite successful, but has several disadvantages:

(i) we often need to work with operators on Banach spaces of C^1 or Hölder functions, despite the smoothness of the diffeomorphism or flow (given by the regularity of the stable foliations);

(ii) this makes it particularly difficult to get a meromorphic extension to $\mathbb{C}$ (because of the existence of the essential spectrum of the operator, as we commented earlier);
(iii) it is very cumbersome to convert invertible systems to noninvertible systems just to introduce some transfer operator (or averaging operator).

Therefore it is desirable to develop a new approach to overcome these. In the classical approach, the invertible system $T : X \to X$, typically the Poincaré map with respect to Markov sections for the geodesic flow, gives rise to a noninvertible system (with local inverses T_i) which gives a transfer operator averaging over the preimages under T_i. However, in the new approach the invertible system is again studied. But now one introduces a Banach space of anisotropic distributions (generalized functions). The transfer operator is essentially simple composition.

10.1. Banach spaces of anisotropic analytic distributions. Historically, the first step was for real analytic Anosov diffeomorphisms, and was initiated by H. Rugh [62], [63]. Recall that we can divide $\mathbb{T}^2$ into elements of a Markov partition $\{\mathcal{T}_i\}$ These have natural real analytic coordinates $(x_i, y_i) \in \mathcal{T}_i$ and let $(x_j, y_j) = T(x_i, y_i) \in \mathcal{T}_j$. Let us write

$$T(x_i, y_i) = (f_1(x_i, y_i), f_2(x_i, y_i)).$$

Let D_i^u, D_i^s be disks in the complexification of the coordinates.

(a) We can solve

$$f_2(x_i, \phi_s(x_i, y_j)) = y_j$$

to get a family of contractions

$$\phi_s(x_i, \cdot) : D_j^u \to D_i^u$$

indexed by x_i.

(b) We can then define a family of contractions $\phi_u(\cdot, y_j) : D_i^s \to D_j^s$ indexed by y_j by

$$\phi_u(x_i, y_j) = f_1(\phi_s(x_i, y_j), y_j).$$

(Note that if f was linear then the foliation would be straight lines and then ϕ_s would also be linear.)

(c) We can define an operator on distributions on $\bigcup_i \mathcal{T}_i$ by

$$\mathcal{L}\psi(x_i, y_i) = \sum_{j:A(i,j)=1} \left(\frac{-1}{2\pi i}\right)^2 \int_{\partial D_j^s} \int_{\partial D_j^u} \int_{\partial D_j^s} \frac{dx_j dy_j \psi(x_j, y_j) \times \partial_2 \phi_s(x_i, y_j)}{(x_i - \phi_u(x_i, y_j))(j_j - \phi_s(x_i, y_j))}$$

defined on the Banach space of analytic functions on $\sum_j (\mathbb{C} - D_j^s) \times D_j^u$.

Remarkably, the operator is nuclear (and thus trace class) and has trace

$$\operatorname{trace}(\mathcal{L}^n) = \sum_{f^n x = x} \frac{1}{|\det(D_x T^n - I)|}.$$

If we choose the coordinates

$$\{z \in \mathbb{C} : |z| > 1\} \times \{w \in \mathbb{C} : |z| < 1\}$$

the elements can be expanded in terms of $z^{-(n+1)} w^m$ where $n, m \geq 0$.

This construction hints at the use of dual spaces, but still has lots of anachronisms (e.g. Markov partitions).

10.2. Banach spaces of anisotropic smooth distributions. More generally, we can consider the Gouëzel-Liverani approach for Anosov diffeomorphisms [28]. The aim is to construct Banach spaces of distributions built so as to have a special form of "duality", which makes the composition with the Anosov diffeomorphism into a contraction.

Let Σ denote the C^∞ embedded leaves of bounded length (of dimension $\dim E^s$) which lie in a C^∞ cone field close to the stable bundles

$$\mathcal{C} = \{v^s \oplus v^u \in E^s_x \oplus E^u_x : \|v^s\| \leq K\|v^u\|\}$$

for some $K > 0$. One can fix $p, q \geq 1$. Let $w : M \to \mathbb{R}$ be C^∞ and let $D^p w$ be the pth order derivative ($p \geq 1$). Let $C^q_0(W) = \{\phi : W \to \mathbb{R}$ be C^q which vanish on $\partial W\}$ for $W \in \Sigma$ then we define a semi-norm by

$$\|w\|^-_{p,q} = \sup_{W\in\Sigma} \sup_{D} \sup_{\phi\in C^q_0(W)} \int_W D^p w \phi d(vol)$$

(a Sobolev-like inner product) and a norm by

$$\|u\|_{p,q} = \sup_{0\leq k\leq p} \|u\|^-_{p,q+k}.$$

We let $B_{p,q}$ be the completion of $C^\infty(M)$ with respect to $\|\cdot\|_{p,q}$ (There was an earlier attempt at constructing such Banach spaces due to Kitaev [35], but it is a little difficult to understand.) The transfer operator acting on this Banach space takes a simple form.

Definition 10.1. The *transfer operator* takes the form $\mathcal{L} : B_{p,q} \to B_{p,q}$ where

$$\mathcal{L}w = \frac{1}{\det(DT) \circ T^{-1}} w \circ T^{-1}$$

i.e., $\int_M wu \circ T d(vol) = \int_M (\mathcal{L}w)u d(vol)$, which corresponds to a change of variables.

We can consider a particularly simple case:

Example 10.2. When $\det(DT) = 1$ then $\mathcal{L}w = w \circ T^{-1}$.

The main result that ultimately leads to a host of applications is the quasi-compactness of this operator (with bounds on the essential spectral radius). The next lemma summarizes the useful spectral properties of $\mathcal{L}$ on this space.

Theorem 10.3. *Let $0 < \theta < 1$ be determined by the expansion and contraction rates. Then*

1. *$\mathcal{L}$ has a maximal positive eigenvalue (and eigenprojection μ corresponding to the SRB measure); and*
2. *$\mathcal{L} : B_{p,q} \to B_{p,q}$ has only isolated eigenvalues in $|z| > \theta^{\min\{p,q\}}$.*

Thus the larger one chooses p, q the more fine structure of the spectrum is revealed. The proof of Theorem 10.3 parallels those in which the quasi-compactness of the earlier transfer operators was established. In particular, it is based on two ingredients, which we formulate in the next two lemmas.

Lemma 10.4. *The unit ball in $B_{p,q} \subset B_{p-1,q+1}$ is relatively compact.*

Lemma 10.5 (Doeblin-Fortet/Lasota-Yorke). *There exist $A, B > 0$ such that for all $n \geq 0$,*

$$\|\mathcal{L}^n w\|_{p,q} \leq A\theta^{\min\{p,q\}n}\|w\|_{p,q} + B\|w\|_{p-1,q+1}$$

Here $\|\cdot\|_{p-1,q+1}$ is the "weak norm" and $\|\cdot\|_{p,q}$ is the "strong norm".

We briefly describe the proof of the Fortet-Doeblin/Lasota-Yorke inequality in Lemma 10.5. To establish this, one needs to estimate terms like

$$\int_W D^k(\mathcal{L}w)\phi d(vol)_W$$

where $0 \leq k \leq p$, $\phi \in C^\infty(W)$. Let us try and explain the basic idea in the construction. Let $n \gg 1$. Then since $T^{-n}W$ is "long" we can break it into standard size pieces: $T^{-n}W = \bigcup_j W_j$. Thus

$$\int_W D^k(\mathcal{L}^n w)\phi d(vol) = \sum_j \int_{T^n W_j} D^k(\mathcal{L}^n w)\phi d(vol).$$

Writing $D = D_u + D_s$, with derivatives D_s along W and D_u "close" to the unstable direction gives terms

$$\int_{T^n W_j} D_s^l D_u^{k-l}(w \circ T^{-n})\phi d(vol) + O(\|w\|_{p-1,q+1})$$

where the error term is the price of reordering with D_s to the front. Integrating by parts moves D_s^l to give

$$\int_{T^n W_j} D_u^{k-l}(w \circ T^{-n}) D_s^l \phi d(vol).$$

By a change of variables (using T^n), we get

$$\int_{W_j} D_u^{k-l}(w) D_s^l \phi \circ T^n d(vol).$$

One contribution comes from $k = p$ and $l = 0$ (the others are dominated by $\|h\|_{p-1,q+1}$). Then

$$\int_{W_j} D_u^p(w)\phi \circ T^n d(vol) = O(\theta^{pn}\|w\|) = O(\theta^{pn}\|w\|_{p,q})$$

where $D_u^p(w)$ contributes the scaling by θ^{pn} and $\phi \circ T^n$ and then we can sum. Note that the contribution from the term $l = k = p$ is $O(\theta^{qn}\|w\|)$.

Remark 10.6. Other Gibbs measures require modifying the norms fundamentally. A more comprehensive discussion of related anisotropic Banach spaces can be found in [6].

10.3. Anosov flows. We want to move from the setting of Anosov diffeomorphisms to that of Anosov flows. To study dynamical properties of Anosov flows we would like to use a similar approach to that for the particular case of geodesic flows. Using the Butterey-Liverani approach for the Anosov flows $\phi_t : M \to M$ we can associate suitable Banach spaces $B_{p,q}$ [19]. The definition of these Banach spaces for Anosov flows is analogous to that for Anosov diffeomorphisms. (We can use Σ to denote a space of C^∞ curves close to the strong stable leaves, i.e., lying in a C^∞ cone family.)

We next define a suitable operator for the Anosov flow.

Definition 10.7. We can define operators $\mathcal{L}_t : B_{p,q} \to B_{p,q}$ $(t > 0)$ by

$$\mathcal{L}_t w = \frac{w \circ \phi_t}{\det(D\phi) \circ \phi_{-t}}$$

and the resolvent operator(s) $R(z) : B_{p,q} \to B_{p,q}$ by

$$R(z)w = \int_0^\infty e^{-zt} \mathcal{L}_t dt$$

for $Re(z) > 0$.

One of the many advantages of the use of anisotropic spaces is that we can work directly with the resolvent operator $R(s)$. This luxury was not previously available to use when we used suspension semi-flows.

The next result describes the meromorphic extension of the resolvent [19]. Let $\lambda > 0$ be the contraction rate for the Anosov flow.

Theorem 10.8. *The operator $R(z)$ is meromorphic for $Re(z) > -\lambda \min\{p, q\}$.*

In particular, recall that given an Anosov flow we can consider the correlation function

$$\rho(t) = \int F \circ \phi_r G d\mu - \int F d\mu \int G d\mu$$

where $F, G \in C^\infty(M)$ and μ is the invariant volume (or more generally the SRB measure). We can deduce the following result on the meromorphic extension of the Laplace transform of the correlation function.

Theorem 10.9. *The Laplace transform*

$$\widehat{\rho}(z) = \int_0^\infty e^{-zt\rho(t)} dt$$

is meromorphic for $Re(z) > -\lambda \min\{p, q\}$ (for all $z \in \mathbb{C}$ if we can choose p, q arbitrarily large).

To study the periodic orbits for the Anosov flow $\phi_t : M \to M$ we can define a zeta function as follows.

Definition 10.10. Given an Anosov flow we can formally define the zeta function

$$\zeta(s) = \prod_\tau (1 - e^{-s\lambda(\tau)})^{-1}, \qquad s \in \mathbb{C},$$

where τ denotes a (primitive) closed orbit of least period $\lambda(\tau)$.

The meromorphic extension of this complex function is again based on the analysis of the transfer operator. By choosing p, q sufficiently large we get

Theorem 10.11. *The zeta function $\zeta(s)$ for a C^∞ Anosov flow is meromorphic for all $s \in \mathbb{C}$. The value $s = h$ is a simple pole for $\zeta(s)$* [30].

We briefly describe the main steps in the proof.

Step 1 (The role of s). Recall that in Definition 10.7 we defined linear operators $R(s) : B_{p,q} \to B_{p,q}$ by

$$R(s)w = \int_0^\infty e^{-st} \mathcal{L}_t dt$$

for $Re(s) > 0$.

Step 2 ("Better" Banach spaces). We replace $C^\infty(M)$ by a Banach space of distributions $B^{(0)}_{p,q}$ and, more generally, construct Banach spaces $B^{(l)}_{p,q}(M)$ for $l = 0, \ldots, \dim M$ replacing functions by l-forms. This gives families of operators: $R^{(l)}_s : B^{(l)}_{p,q} \to B^{(l)}_{p,q}$ defined by analogy to $R^{(0)}_s$.

For simplicity, consider $\dim M = 3$ and denote $\sigma_1 = h$, where h is the topological entropy of the flow and $\sigma_0 = \sigma_2 = h - \sigma$ where $\lambda > 0$ is a bound on the exponential contraction.

Proposition 10.12 (Spectrum of $R^{(l)}_s : B^{(l)}_{p,q} \to B^{(l)}_{p,q}$). *Assume that* $Re(s) > \sigma_l$ $(l = 0, 1, 2)$. *Then*

(a) *the spectral radius satisfies* $\rho(R^{(l)}_s) \le \frac{1}{Re(s)-\sigma_l}$;
(b) *the essential spectral radius satisfies*

$$\rho_e(R^{(l)}_s) \le \frac{1}{Re(s) - \sigma_l + \lambda[(k-2)/2]}$$

where $k = \min\{p, q\}$.

Step 3 (The extension). We can associate to the resolvent a complex function ("the determinant") defined as follows:

$$D_l(s) = \exp\left(- \sum_{n=1}^{\infty} \frac{1}{n} \text{"trace"}(R^{(l)}_s) \right)$$

where we ignore the nonessential part of the spectrum. In particular, $D_l(s)$ is analytic for $Re(s) > \sigma_l - \lambda[(k-2)/2]$. We can then write

$$\zeta(s) = D_0(s) D_2(s) / D_1(s)$$

where the numerator gives zeros for $Re(s) < h - \lambda$. The denominator gives poles for $Re(s) < h$.

In particular, the conclusion is that for C^k Anosov flows, the zeta function $\zeta(s)$ is meromorphic for $Re(s) > h - \lambda[(k-2)/2]$, and letting $k \to +\infty$ gives a meromorphic extension to $\mathbb{C}$.

Remark 10.13. Previous results in the direction include:

(a) Ruelle showed Corollary 10.11 under the additional assumption that the stable manifolds are C^ω [58].
(b) Fried (adapting Rugh's approach) showed the result assuming the flow is C^ω [27].

Remark 10.14. There is another construction of Banach spaces by Dyatlov-Zworski using microlocal analysis [23].

For some geodesic flows there is also an analytic extension to a strip [30]. Let $\phi_t : M \to M$ be the geodesic flow for a compact manifold V with negative sectional curvatures.

Theorem 10.15 ([30]). *For 1/9-pinched negative sectional curvatures, for all* $\epsilon > 0$, $\zeta(s)$ *has a nonzero analytic extension to* $h - \epsilon < Re(s) < h$.

This leads to the following estimate on the number of closed orbits of period at most T.

COROLLARY 10.16. *For 1/9-pinched negative sectional curvatures,*

$$\mathrm{Card}\{\tau : \lambda(\tau \leq T)\} = \mathrm{li}(e^{hT})(1 + O(e^{-\epsilon T})).$$

REMARK 10.17. Previous results in the direction include:

(a) This is true for surfaces without extra conditions [52].
(b) The principal term is true for manifolds without the pinching condition [41]:

$$\mathrm{Card}\{\tau : \lambda(\tau \leq T)\} \sim \frac{e^{hT}}{hT}$$

as $T \to +\infty$.

This generalizes to contact Anosov flows with 1/3-pinching.

REMARK 10.18. We can also use this formalism to consider decay of correlations for the maximal entropy measure (or Bowen-Margulis measure) rather than the SRB-measure [30]. Let μ denote the measure of maximal entropy for $\phi_t : M \to M$ and let $F, G \in C^\infty(M)$. Let

$$\rho(t) = \int F \circ \phi_t G d\mu - \int F d\mu \int G d\mu,$$

for $t > 0$, be the correlation functions. The asymptotic behaviour of $\rho(t)$ is given by the analytic properties of the Laplace transform

$$\widehat{\rho}(s) = \int_0^\infty e^{-st}\rho(t)dt, \qquad s \in \mathbb{C}.$$

We observe that:

(a) $\widehat{\rho}(s)$ converges for $Re(s) > 0$;
(b) $\widehat{\rho}(s)$ has a meromorphic extension to $\mathbb{C}$;
(c) typically s is a pole for $\widehat{\rho}(s)$ if $s + h$ is a pole for $\zeta(s)$ (actually zero for $Z(s)$), since both can be related to properties of $R(s)$;
(d) there exist $C > 0, \lambda > 0$ such that $|\rho(t)| \leq Ce^{-\lambda t}$, $t > 0$ providing the curvature is 1/9-pinched.

11. Other notes. The more discerning reader may prefer other notes which have a more specific focus on particular topics.

1. For the reader wanting a more pure and undiluted theory of dynamical zeta functions the author has some unpublished notes from lectures in Grenoble [51] (about 35 pages).
2. For the reader wanting more details on the connections with fractals the author has some notes from lectures in Porto [48] (about 106 pages).
3. For the reader wanting a more geometrical or number theoretical viewpoint, I would recommend reading elsewhere on the Selberg zeta function, e.g., [31].

Acknowledgments. I am very grateful to Richard Sharp and the anonymous referee for reading these notes and helping me to eliminate (some of) the mistakes.

This work was partially supported by the grant 346300 for IMPAN from the Simons Foundation and the matching 2015-2019 Polish MNiSW fund. I am grateful to IMPAN for their hospitality during my stay.

References

[1] R. Adler and B. Weiss, *Similarity of automorphisms of the torus*, Mem. Amer. Math. Soc. 98 (1970).

[2] D. Anosov, *Geodesic flows on closed Riemannian manifolds of negative curvature*, Trudy Mat. Inst. Steklov. 90 (1967) (in Russian).

[3] V. Arnol'd and A. Avez, *Ergodic Problems of Classical Mechanics*, W. A. Benjamin, New York, 1968.

[4] M. Artin and B. Mazur, *On periodic points*, Ann. of Math. (2) 81 (1965), 82-99.

[5] A. Avila, S. Gouëuzel, and J.-C. Yoccoz, *Exponential mixing for the Teichmüller flow*, Publ. Math. Inst. Hautes Études Sci. No. 104 (2006), 143–211.

[6] V. Baladi, *Dynamical zeta functions and dynamical determinants for hyperbolic maps*, preprint.

[7] V. Baladi and B. Vallée, *Euclidean algorithms are Gaussian*, J. Number Theory 110 (2005), 331–386.

[8] O. Bandtlow and O. Jenkinson, *Explicit eigenvalue estimates for transfer operators acting on spaces of holomorphic functions*, Adv. Math. 218 (2008), 902–925.

[9] D. Borthwick, *Distribution of resonances for hyperbolic surfaces*, Exp. Math. 23 (2014), 25–45.

[10] J. Bourgain, A. Gamburd, and P. Sarnak, *Generalization of Selberg's $\frac{3}{16}$ theorem and affine sieve*, Acta Math. 207 (2011), 255–290.

[11] J. Bourgain and A. Kontorovich, *On Zaremba's conjecture*, Ann. of Math. (2) 180 (2014), 137–196.

[12] R. Bowen, *Hausdorff dimension of quasicircles*, Inst. Hautes Études Sci. Publ. Math. 50 (1979), 11–25.

[13] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Math. 470, Springer, Berlin, 1975.

[14] R. Bowen and D. Ruelle, *The ergodic theory of Axiom A flows*, Invent. Math. 29 (1975), 181–202.

[15] R. Bowen and C. Series, *Markov maps associated with Fuchsian groups*, Inst. Hautes Études Sci. Publ. Math. 50 (1979), 153–170.

[16] M. Bridgeman, R. Canary, F. Labourie, and A. Sambarino, *The pressure metric for Anosov representations*, Geom. Funct. Anal. 25 (2015), 1089–1179.

[17] K. Burns, H. Masur, C. Matheus, and A. Wilkinson, *Rates of mixing for the Weil-Petersson geodesic flow: exponential mixing in exceptional moduli spaces*, Geom. Funct. Anal. 27 (2017), 240–288.

[18] K. Burns, H. Masur, and A. Wilkinson, *The Weil-Petersson geodesic flow is ergodic*, Ann. of Math. 175 (2012), 835–908.

[19] O. Butterley and C. Liverani, *Smooth Anosov flows: correlation spectra and stability*, J. Mod. Dyn. 1 (2007), 301–322.

[20] H. Davenport, *Multiplicative Number Theory*, 2nd ed., Grad. Texts in Math. 74, Springer, New York, 1980.

[21] D. Dolgopyat, *On decay of correlations in Anosov flows*, Ann. of Math. 147 (1998), 357–390.

[22] D. Dolgopyat, *Prevalence of rapid mixing in hyperbolic flows*, Ergod. Th. Dynam. Sys. 18 (1998), 1097–1114.

[23] S. Dyatlov and M. Zworski, *Dynamical zeta functions for Anosov flows via microlocal analysis*, Ann. Sci. École Norm. Sup. 49 (2016), 543–577.

[24] W. Ellison, *Les nombres premiers*, Hermann, Paris, 1975.

[25] K. Falconer, *Fractal Geometry, Mathematical Foundations and Applications*, Wiley, Chichester, 1990.

[26] D. Fried, *The zeta functions of Ruelle and Selberg*, I. Ann. Sci. École Norm. Sup. (4) 19 (1986), 491–517.

[27] D. Fried, *Meromorphic zeta functions for analytic flows*, Comm. Math. Phys. 174 (1995), 161–190.

[28] S. Gouëzel and C. Liverani, *Banach spaces adapted to Anosov systems*, Ergod. Th. Dynam. Sys. 26 (2006), 189–217.

[29] A. Grothendieck, *Produits tensoriels topologiques et espaces nucléaires*, Mem. Amer. Math. Soc. 16 (1955).

[30] P. Giulietti, C. Liverani, and M. Pollicott, *Anosov flows and dynamical zeta functions*, Ann. of Math. 178 (2013), 687–773.

[31] D. Hejhal, *The Selberg Trace Formula for PSL(2,R)*, Vol. I, Lecture Notes in Math. 548, Springer, Berlin, 1976.

[32] M. Hirsch and C. Pugh, *Smoothness of horocycle foliations*, J. Differential Geometry 10 (1975), 225–238.

[33] O. Jenkinson and M. Pollicott, *Computing invariant densities and metric entropy*, Comm. Math. Phys. 211 (2000), 687–703.

[34] O. Jenkinson and M. Pollicott, *Rigorous effective bounds on the Hausdorff dimension of continued fraction Cantor sets: a hundred decimal digits for the dimension of E2*, preprint.

[35] A. Kitaev, *Fredholm determinants for hyperbolic diffeomorphisms of finite smoothness*, Nonlinearity 12 (1999), 141–179.

[36] A. Kontorovich, *From Apollonius to Zaremba: local-global phenomena in thin orbits*, Bull. Amer. Math. Soc. (N.S.) 50 (2013), 187–228.

[37] A. Manning, *Axiom A diffeomorphisms have rational zeta functions*, Bull. London Math. Soc. 3 (1971), 215–220.

[38] A. Kontorovich and H. Oh, *Apollonian circle packings and closed horospheres on hyperbolic 3-manifolds* (with an appendix by H. Oh and N. Shah), J. Amer. Math. Soc. 24 (2011), 603–648.

[39] F. Labourie, *Anosov flows, surface groups and curves in projective space*, Invent. Math. 165 (2006), 51–114.

[40] C. Liverani, *On contact Anosov flows*, Ann. of Math. (2) 159 (2004), 1275–1312.

[41] G. Margulis, *Certain applications of ergodic theory to the investigation of manifolds of negative curvature*, Funktsional. Anal. i Prilozhen. 3 (1969), no. 4, 89–90 (in Russian).

[42] G. Margulis, *On Some Aspects of the Theory of Anosov Systems*, Springer Monographs in Mathematics, Springer, Berlin, 2004.

[43] R. Mauldin and M. Urbański, *Dimension and measures for a curvilinear Sierpiński gasket or Apollonian packing*, Adv. Math. 136 (1998), 26–38.

[44] D. Mayer, *On a ζ function related to the continued fraction transformation*, Bull. Soc. Math. France 104 (1976), 195–203.

[45] C. Moreira, *Fractal geometry and dynamical bifurcations*, in: Proc. ICM (Seoul, 2014), Vol. III, 647–659.

[46] C. McMullen, *Thermodynamics, dimension and the Weil-Petersson metric*, Invent. Math. 173 (2008), 365–425.

[47] W. Parry and M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque 187-188 (1990), 1–268.

[48] M. Pollicott, *Fractals and Dimension Theory*, introductory lectures given in Porto 2005, homepages.warwick.ac.uk/∼masdbl/dimensional-total.pdf.

[49] M. Pollicott, *A complex Ruelle-Perron-Frobenius theorem and two counterexamples*, Ergod. Th. Dynam. Sys. 4 (1984), 135–146.

[50] M. Pollicott, *On the rate of mixing of Axiom A flows*, Invent. Math. 81 (1985), 413–426.

[51] M. Pollicott, *Dynamical zeta functions*, uncorrected notes of lectures given in Grenoble (2013), homepages.warwick.ac.uk/∼masdbl/grenoble-16july.pdf.

[52] M. Pollicott and R. Sharp, *Exponential error terms for growth functions on negatively curved surfaces*, Amer. J. Math. 120 (1998), 1019–1042.

[53] M. Pollicott, *Some applications of thermodynamic formalism to manifolds with constant negative curvature*, Adv. Math. 85 (1991), 161–192.

[54] M. Pollicott and M. Urbański, *Asymptotic counting in conformal dynamical systems*, Mem. Amer. Math. Soc., to appear.

[55] M. Pollicott and P. Vytnova, *Zeros of the Selberg zeta function for non-compact surfaces*, preprint.

[56] M. Ratner, *Markov decomposition for an U-flow on a three-dimensional manifold*, Mat. Zametki 6 (1969), 693–704.

[57] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. I. Functional Analysis*, Academic Press, New York, 1980

[58] D. Ruelle, *Zeta-functions for expanding maps and Anosov flows*, Invent. Math. 34 (1976), 231–242.

[59] D. Ruelle, *Thermodynamic Formalism*, Encyclopedia Math. Appl. 5, Addison-Wesley, Reading, MA, 1978.

[60] D. Ruelle, *Flots qui ne mélangent pas exponentiellement*, C. R. Acad. Sci. Paris Sér. I Math. 296 (1983), 191–193.

[61] D. Ruelle, *Repellers for real analytic maps*, Ergod. Th. Dynam. Sys. 2 (1982), 99–107.

[62] H. Rugh, *The correlation spectrum for hyperbolic analytic maps*, Nonlinearity 5 (1992), 1237–1263.

[63] H. Rugh, *Generalized Fredholm determinants and Selberg zeta functions for Axiom A dynamical systems*, Ergod. Th. Dynam. Sys. 16 (1996), 805–819.

[64] C. Series, *Geometrical Markov coding of geodesics on surfaces of constant negative curvature*, Ergod. Th. Dynam. Sys. 6 (1986), 601–625.

[65] S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. 73 (1967), 747–817.