Dynamical Zeta Functions and Transfer Operators

David Ruelle

Certain generating functions—encoding properties of objects like prime numbers, periodic orbits, …—have received the name of *zeta functions*. They are useful in studying the statistical properties of the objects in question. Zeta functions have generally been associated with problems of arithmetic or algebra and tend to have common features: meromorphy, Euler product formula, functional equation, location of poles and zeros (Dirichlet series expansion, Riemann hypothesis), and relation with certain operators (typically operators acting on cohomology groups). The *dynamical zeta functions* to be discussed here are set up to count periodic orbits but to count them with fairly general *weights*. As a consequence the subject will have a more function-theoretic flavor than the study of arithmetic or algebraic zeta functions. Apart from that, our zeta functions will have properties similar to those of the more traditional ones. The main difference will be that the relevant operators (called *transfer operators*) will act on (infinite-dimensional) cochain groups instead of (finite-dimensional) cohomology groups. Intuitively, the weights that we have introduced prevent passage from cochains to cohomology groups. Technically this will force us to consider determinants in infinite dimension. The study of dynamical zeta functions uses original tools (transfer operators, kneading determinants), which we shall discuss below.

The simplest invariant measures for a dynamical system are those carried by periodic orbits. Counting periodic orbits is thus a natural task from the point of view of ergodic theory. And dynamical zeta functions are an effective tool to do the counting. The tool turns out to be so effective in fact as to make one suspect that there is more to the story than what we currently understand.

David Ruelle is professor of mathematics at Rutgers University and the Institut des Hautes Études Scientifiques, Bures-sur-Yvette, France. His e-mail address is ruelle@ihes.fr.

Some Traditional Examples of Zeta Functions

The grandmother of all zeta functions is the Riemann zeta function defined by

$$\zeta_R(s) = \sum_{n=1}^{\infty} n^{-s}$$

for Re $s > 1$. Actually, this function was first considered by Euler, who noted that

$$\zeta_R(s) = \prod_{p \text{ prime}} (1 - p^{-s})^{-1}$$

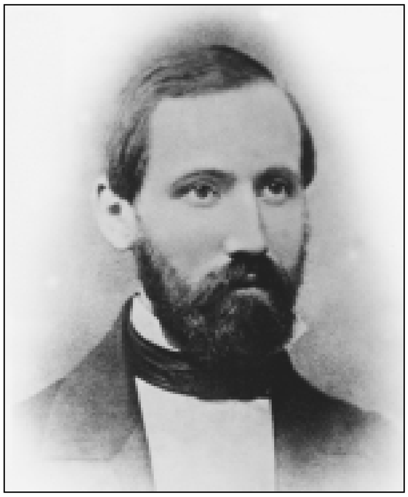
(this is the *Euler product formula*). Riemann showed that $\zeta_R(s)$ extends meromorphically to $\mathbf{C}$ with a single pole at $s = 1$ and that there is a *functional equation* relating $\zeta_R(s)$ and $\zeta_R(1-s)$. Because ζ_R is a generating function for the primes, it can be used to prove the prime number theorem: that the number of primes up to x is $\sim x/\log x$. A theorem from harmonic analysis called the Wiener-Ikehara Tauberian theorem yields the prime number theorem from the fact that $\zeta_R(s)$ has a simple pole at $s = 1$ and no other pole or zero for Re $s \geq 1$.

After the Riemann zeta function, innumerable functions with related properties have been introduced. In particular, given an algebraic variety over a finite field $\mathbf{F}_q$, we may define a "Weil zeta function" by

$$\zeta_W(z) = \exp \sum_{m=1}^{\infty} \frac{z^m}{m} |\mathrm{Fix} f^m|.$$

Here one has extended the algebraic variety to the algebraic closure of $\mathbf{F}_q$, obtaining a space M, and $f : M \to M$ is the Frobenius map (acting by $z \mapsto z^q$ on coordinates); $|\mathrm{Fix}\, f^m|$ is the number of fixed points of the m-th iterate of f. The function $\zeta_W(z)$ satisfies the Weil conjectures (Weil, Dwork, Grothendieck, Deligne); in particular it is rational. Note

AIP Emilio Segrè Visual Archives, T. J. J. See Collection.



The idea of using the zeta function to study the asymptotic distribution of primes is due to Georg Friedrich Bernhard Riemann (1826-1866), perhaps the greatest mathematician of all times.

that the variable z in $\zeta_W(z)$ has to be thought of as the exponential of $-s$ in $\zeta_R(s)$.

The Weil zeta function counts periodic points (or periodic orbits) for the dynamical system (M, f), where f is the Frobenius map. It is natural to consider a more general space M and map $f : M \to M$ and (assuming that $|\mathrm{Fix}\, f^m|$ is finite for each m) to define

$$\zeta(z) = \exp \sum_{m=1}^{\infty} \frac{z^m}{m} |\mathrm{Fix} f^m|.$$

We have here again an "Euler product formula", namely, the following identity between formal power series:

$$\zeta(z) = \prod_P (1 - z^{|P|})^{-1} \tag{1}$$

where the product is over periodic orbits P and $|P|$ is the period of P. For example, one can take for f a diffeomorphism of a compact manifold M (Artin-Mazur). In the special case when f is hyperbolic (technically, f is an Axiom A diffeomorphism restricted to a basic set), one finds that this zeta function is rational (Smale, Guckenheimer, Manning, Bowen, Fried).

As an example of (1), consider the map $x \mapsto 1 - \mu x^2$ of the interval $[-1, 1]$ to itself. For the Feigenbaum value $\mu = 1.401155\ldots$, this map has one periodic orbit of period 2^n for each integer $n \geq 0$. Therefore

$$\zeta(z) = \prod_{n=0}^{\infty} (1 - z^{2^n})^{-1} = \prod_{n=0}^{\infty} (1 + z^{2^n})^{n+1}$$

where we have used (1) and $(1 - z)^{-1} = \prod_{n=0}^{\infty} (1 + z^{2^n})$. Note that this ζ satisfies the functional equation $\zeta(z^2) = (1 - z)\zeta(z)$.

A natural way to count periodic orbits for a map f is to weight them with the topological index $L(x, f)$. Specifically, assume that f is a diffeomorphism of the compact manifold M, $x \in \mathrm{Fix} f$, and $1 - T_x f$ is invertible (where $T_x f$ is the tangent map to f at x). Then

$$L(x, f) = \mathrm{sgn}\det(1 - T_x f)$$

and we have the Lefschetz trace formula

$$\sum_{k=0}^{\dim M} (-1)^k \mathrm{tr}\, f_{*k} = \sum_{x \in \mathrm{Fix} f} L(x, f)$$

where f_{*k} is the action of f on the k-th homology group of M with real coefficients. Suppose now that $1 - T_x f^m$ is invertible for all fixed points x of f^m for all $m > 0$, and define the Lefschetz zeta function

$$\zeta_L(z) = \exp \sum_{m=1}^{\infty} \frac{z^m}{m} \sum_{x \in \mathrm{Fix} f^m} L(x, f^m).$$

Then the trace formula yields

$$\zeta_L(z) = \prod_{k=0}^{\dim M} \det(1 - z f_{*k})^{(-1)^{k+1}}$$

(therefore $\zeta_L(z)$ is rational). In many interesting cases $L(x, f^m) = 1$ for all periodic points x, so that $\zeta_L(z) = \zeta(z)$.

Suppose now that instead of a discrete time dynamical system generated by $f : M \to M$, we have a continuous time dynamical system, i.e., a semiflow or flow (f^t) on M. Then the Euler formula (1) with z replaced by e^{-s} suggests defining a zeta function

$$\zeta(s) = \prod_{\varpi} (1 - e^{-s\ell(\varpi)})^{-1} \tag{2}$$

where the product is over (prime) periodic orbits ϖ and $\ell(\varpi)$ is the period of ϖ.

A much-studied example of a flow is the geodesic flow on a Riemann manifold N. We recall the definition. If a point $x(t)$ moves at unit speed along a geodesic of N and $u(t) = \frac{d}{dt} x(t) \in T_{x(t)} N$, we have $\|u(t)\| = 1$. Writing $u(t) = f^t u(0)$ defines a diffeomorphism f^t of the unit tangent bundle M of N, and $(f^t)_{t \in \mathbf{R}}$ is called the geodesic flow. Note that it is a flow on the unit tangent bundle M rather than on N. Observe also that the period of a periodic orbit for the geodesic flow is the length of a closed geodesic on N.

Closely related to (2) is the definition of the Selberg zeta function ζ_S. This zeta function appears in questions of arithmetic and is defined in terms of a Fuchsian group $\Gamma \subset SL(2, \mathbf{R})$ operating on the complex upper half-plane H (and a matrix representation of Γ which we shall ignore here). If Γ is torsion-free and $\Gamma \backslash H$ is compact, then $\Gamma \backslash H$ is a compact surface with curvature -1 (because H with the Poincaré metric is the Lobatchevsky plane with curvature -1, the geodesics of H are half-circles centered on the real axis). Let (f^t) be the geodesic flow on $\Gamma\backslash H$ so that the periods $\ell(\varpi)$ in (2) are the lengths of the closed geodesics. Then the Selberg zeta function is

$$\zeta_S(s) = \prod_{k=0}^{\infty} \zeta(s+k)^{-1} = \prod_{\varpi} \prod_{k=0}^{\infty} (1 - [\exp \ell(\varpi)]^{-s-k}).$$

It can be shown that ζ_S is an entire analytic function satisfying a functional equation and a form of the Riemann hypothesis. In fact the zeros of ζ_S are related to the eigenvalues of the Laplace-Beltrami operator Δ on $\Gamma\backslash H$. We thus have a connection between classical mechanics (the geodesic flow) and quantum mechanics (with the Hamiltonian Δ). This connection has been much studied in relation with "quantum chaos".

To conclude our list of examples let us mention the currently popular Ihara-Selberg zeta function associated with a finite unoriented graph G. This function ζ_I is of the form (1) where periodic orbits are replaced by cycles (circuits on G without immediate backtracking). It is known that $1/\zeta_I$ is a polynomial and that ζ_I satisfies the Riemann hypothesis precisely when G is Ramanujan (Ramanujan graphs were named by Lubotzky, Phillips, and Sarnak; examples are not easy to construct).

Dynamical Zeta Functions

Let us now equip the dynamical system (M, f), where f need not be invertible but $\mathrm{Fix} f^m$ is finite for all $m > 0$, with a weight $g : M \to \mathbf{C}$ (real positive weights will be of special interest). A zeta function associated with the weighted dynamical system (M, f, g) is defined by

$$\zeta(z) = \exp \sum_{m=1}^{\infty} \frac{z^m}{m} \sum_{x \in \mathrm{Fix} f^m} \prod_{k=0}^{m-1} g(f^k x) \tag{3}$$

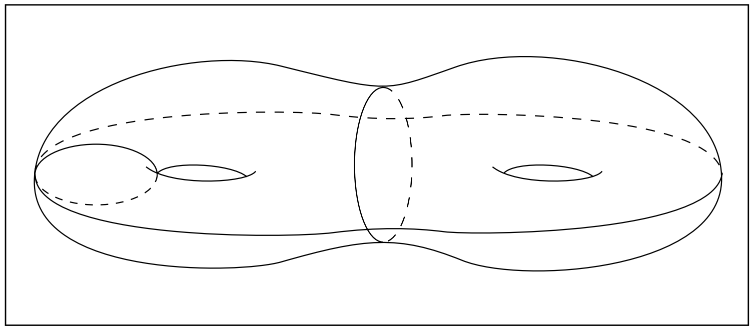
as a formal power series. This is the prototype of what we want to call a dynamical zeta function. We have here again an Euler product formula

$$\zeta(z) = \prod_P \left(1 - z^{|P|} \prod_{k=0}^{|P|-1} g(f^k x_P)\right)^{-1}$$

where $x_P \in P$ is chosen arbitrarily. So, introducing a weight does not spoil the basic combinatorial properties of the zeta function.

What about analyticity? Can we get more analyticity than is immediately obvious and then make use of it to obtain statistical properties of the (weighted) periodic orbits we are counting here? To be specific, suppose that $g = \exp A$, where A is a real function, and write

$$P(A) = \limsup_{m \to \infty} \frac{1}{m} \log \sum_{x \in \mathrm{Fix} f^m} \exp \sum_{k=0}^{m-1} A(f^k x).$$



A compact surface on which a metric with curvature -1 can be put: several geodesics have been drawn.

Then the radius of convergence of $\zeta(z)$ is $\exp(-P(A))$. Can we prove more: that $\zeta(z)$ has an isolated pole at $\exp(-P(A))$? This could, for instance, be used to prove analyticity of $A \mapsto P(A)$. We shall now give an example of this situation.

Let t be an $r \times r$ matrix with elements $t_{ij} = 0$ or 1. Define

$$\Omega = \{(\xi_k)_{k \in \mathbf{Z}} : t_{\xi_k \xi_{k+1}} = 1 \text{ for all } k\}$$

and let $d((\xi_k), (\eta_k)) = \exp[-\min\{|k| : \xi_k \neq \eta_k\}]$. Then Ω is a compact metric space with respect to the metric d. The map $\tau : (\xi_k) \mapsto (\xi_{k+1})$ is called a *shift* and is a homeomorphism of Ω. If we assume that $r \geq 2$ and that for some power t^N of t all the matrix elements t^N_{ij} are positive, the dynamical system (Ω, τ) is called a mixing *subshift of finite type*. Let $C^\alpha(\Omega)$ be the Banach space of real α-Hölder continuous functions on Ω. Since the α-Hölder norm is given by

$$\|A\|_\alpha = \max\{\sup_\xi |A(\xi)|, \sup_{\xi \neq \eta} \frac{|A(\xi) - A(\eta)|}{d(\xi, \eta)^\alpha}\},$$

we see that $A \in C^\alpha(\Omega)$ says that the dependence of $A((\xi_k)_{k \in \mathbf{Z}})$ on ξ_k is exponentially small for large k (bounded by $\|A\|_\alpha e^{-\alpha|k|}$).

1. *Proposition.* *The limit*

$$P(A) = \lim_{m \to \infty} \frac{1}{m} \log \sum_{\xi \in \mathrm{Fix} \tau^m} \exp \sum_{k=0}^{m-1} A(\tau^k \xi) \tag{4}$$

exists, and there is $\mathcal{R} > \exp(-P(A))$ such that the dynamical zeta function $\zeta(z)$ associated with the weighted dynamical system $(\Omega, \tau, \exp A)$ is meromorphic for $|z| < \mathcal{R}$, with a single pole at $\exp(-P(A))$ and no other pole or zero.

[Note that if $A = 0$, the zeta function counts periodic orbits with weight 1 and can be computed exactly (Bowen-Lanford) because $|\mathrm{Fix} \tau^m| = \mathrm{tr}\ t^m$, as one readily checks. Here one finds

$$\zeta(z) = \exp \sum_{m=1}^{\infty} \frac{z^m}{m} \mathrm{tr}\, t^m$$
$$= \exp(-\mathrm{tr}\ \log(1 - zt)) = 1/\det(1 - zt).]$$

The function $A \to P(A)$, called *pressure*, arises in a theory called *thermodynamic formalism* which is based on ideas and methods of statistical mechanics. Having obtained the above nontrivial but apparently useless result, I put it as Exercise 7(c) on page 101 in my book *Thermodynamic Formalism* [3]. A few years later (December 29, 1982) Bill Parry of Warwick wrote to me about very interesting results on Axiom A flows he had obtained with his student Mark Pollicott. These results used Exercise 7(c), which unfortunately he had been unable to do. Could I help? By the time I had (painfully) managed to reconstruct the solution of the exercise I received another letter: *13 Jan 83 / Dear David, / We've finally managed to do your exercise! So ignore my last letter. / Sincerely / Bill Parry.*

Before we look into the work of Parry and Pollicott, let me remark on a relation between the zeta function (2) for a flow and the dynamical zeta function (3). Let M be a compact manifold, $f : M \to M$ a diffeomorphism, and $T : M \to \mathbf{R}$ a smooth positive function. A manifold $\widetilde{M}$ is obtained by identifying in $\{(x, \tau) : x \in M,\ 0 \le \tau \le T(x)\}$ the points $(x, T(x))$ and $(fx, 0)$. Furthermore, there is a smooth flow $(\tilde{f}^t)$ on $\widetilde{M}$ such that $\tilde{f}^t(x,\tau) = (x, t+\tau)$ if $0 \le \tau + t \le T(x)$. This flow $(\tilde{f}^t)$ is called the suspension of f corresponding to the ceiling function T. It is now easy to check that the zeta function $\tilde{\zeta}$ defined by (2) for the flow $(\tilde{f}^t)$ satisfies

$$\tilde{\zeta}(s) = \exp \sum_{m=1}^{\infty} \frac{1}{m} \sum_{x \in \mathrm{Fix} f^m} \prod_{k=0}^{m-1} e^{-sT(f^k x)}$$

and is thus equal to the dynamical zeta function (3) for $z = 1$ and $g = e^{-sT}$. In particular, $\tilde{\zeta}(s)$ will be analytic in s when $\zeta(z)$ defined by (3) is analytic at $z = 1$.

By the way, it is natural to introduce a generalization of (2) associated with a function $B : M \to \mathbf{C}$, viz.

$$\zeta(s) = \prod_{\varpi} \Big(1 - \exp\Big[-s \int_0^{\ell(\varpi)} dt\, B(f^t x(\varpi))\Big]\Big)^{-1},$$

where $x(\varpi)$ is an arbitrarily chosen point in ϖ. In the case of a suspension this is again related to (3).

Hyperbolic Dynamics and Thermodynamic Formalism

Let K be a compact invariant set for the C^r diffeomorphism $f : M \to M$. One says that K is hyperbolic if the tangent bundle restricted to K has a continuous splitting

$$T_K M = V^s \oplus V^u$$

invariant under Tf and such that, for a suitable Riemann metric and $0 < \theta < 1$,

$$\|T_x f^n v\| \le \|v\| \theta^n \quad \text{when} \quad v \in V^s_x, n \ge 0$$
$$\|T_x f^{-n} v\| \le \|v\| \theta^n \quad \text{when} \quad v \in V^u_x, n \ge 0.$$

If the entire manifold M is hyperbolic, f is called an Anosov diffeomorphism. Of particular interest are the *hyperbolic sets with local product structure*. We shall not define this concept, introduced by Smale, but mention as an example the closure K of the set of periodic points for an Anosov diffeomorphism. (It is conjectured that for an Anosov diffeomorphism, the closure of the set of periodic points is in fact always M itself.)

Arnold's *cat map* is an example of an Anosov diffeomorphism on $\mathbf{R}^2/\mathbf{Z}^2$. It is defined by $(x, y) \mapsto (x + y, x + 2y) \pmod 1$. Hyperbolicity is seen by checking that the eigenvalues of $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ have modulus $\ne 1$.

If K is hyperbolic with local product structure, then f restricted to K is essentially a subshift of finite type. This follows from the existence of *Markov partitions* first proved (after an example of Adler and Weiss) by Sinai for Anosov diffeomorphisms, then by Bowen in the general case. More precisely, there is a subshift of finite type (Ω, τ) and a Hölder continuous map π of Ω onto K such that

a. $f \circ \pi = \pi \circ \tau$,

b. π^{-1} is uniquely defined on a residual set (countable intersection of dense open subsets of K)

c. $\max_{x \in K} |\pi^{-1} x|$ is finite.

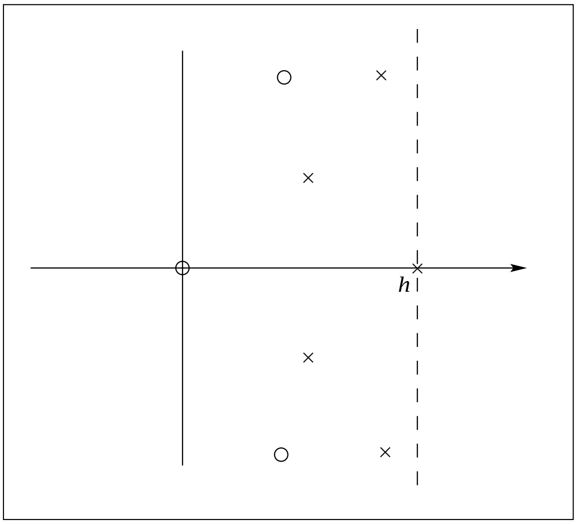
Counting periodic points for τ is not quite the same thing as counting periodic points for f, but almost (Manning and Bowen have shown how to do an exact counting, using finitely many subshifts of finite type). It simplifies matters to assume that f is topologically mixing on K: if O_1, O_2 are nonempty open subsets of K, then $O_1 \cap f^{-n} O_2 \ne \emptyset$ for sufficiently large n. In brief, from Proposition 1 one obtains the following.

2. *Proposition.* *Let K be hyperbolic with local product structure such that $f|K$ is topologically mixing. If $A : K \to \mathbf{R}$ is Hölder continuous, then*

$$P(A) = \lim_{m \to \infty} \frac{1}{m} \log \sum_{x \in \mathrm{Fix} f^m} \exp \sum_{k=0}^{m-1} A(f^k x)$$

exists and there is $\mathcal{R} > \exp(-P(A))$ such that the dynamical zeta function $\zeta(z)$ associated with the weighted dynamical system $(K, f|K, \exp A)$ is meromorphic for $|z| < \mathcal{R}$, with a single pole at $\exp(-P(A))$ and no other pole or zero.

Let me return to the letter that Bill Parry sent me at the end of 1982. His interest was in hyperbolic flows (in particular Anosov flows), which have a theory very similar to hyperbolic (and Anosov) diffeomorphisms. We shall not give detailed definitions but note an important example: *the geodesic flow on a compact manifold of variable negative*



The function $\zeta(s)$ defined by (2) for the geodesic flow on a compact manifold of variable negative curvature has a pole at $h > 0$. Other zeros and poles are in $\{z : \mathrm{Re}z < h\}$.

curvature is an Anosov flow. Bowen has shown that if (f^t) is a smooth flow restricted to a hyperbolic set K with local product structure, then counting periodic orbits for (f^t) is basically the same thing as counting periodic orbits with weights for a subshift of finite type. (This is because (f^t) has a Markov partition; i.e., it is basically a suspension of a subshift of finite type with a suitable ceiling function.) Assuming that (f^t) is topologically mixing, one can then show that $\zeta(s)$ defined by (2) has a meromorphic extension to an open set containing $\{s : \mathrm{Re}\ s \geq h\}$, without zero and with a single pole at $s = h$. (The number $h > 0$ is known as "topological entropy of (f^t) restricted to K", with a general definition that need not concern us here.) The analyticity of $\zeta(s)$ is thus very similar to the analyticity of the Riemann zeta function as used to prove the prime number theorem. The same method (Wiener-Ikehara Tauberian theorem) allowed Parry and Pollicott to prove that the number of periodic orbits ϖ with period $\ell(\varpi) \leq x$ is $\sim e^{hx}/hx$. This extended an earlier result of Margulis by a new and very elegant method. Later Lalley, Katsuda and Sunada, Parry, Pollicott, and Sharp followed the same line of thought and studied the distribution of periods for periodic orbits satisfying various conditions, with error terms, etc. When (f^t) is the geodesic flow on a manifold of variable negative curvature, one thus obtains detailed information about the lengths of geodesics on the manifold. The special case of surfaces of constant curvature -1 is of arithmetic interest (as we said when we introduced Selberg's zeta function). So the study of dynamical zeta functions extends to manifolds of variable negative curvature some results of arithmetic interest known in the case of constant negative curvature.

The Method of Transfer Operators

The proof of Proposition 1 uses *transfer operators*. Given a set Λ (which need not be a manifold) and maps $F : \Lambda \to \Lambda$, $g : \Lambda \to \mathbf{C}$, a transfer operator $\mathcal{L}$ acting on functions $\Phi : \Lambda \to \mathbf{C}$ is defined by

$$(\mathcal{L}\Phi)(x) = \sum_{y \in F^{-1}\{x\}} g(y)\Phi(y)\,. \tag{5}$$

[As an example, if F has Jacobian determinant J and $g = 1/|J|$, the direct image by F of the measure $\Phi(x)dx$ is $(\mathcal{L}\Phi)(x)dx$.] The situation to keep in mind is when F is finite-to-one, expanding, and the functions g and Φ have some kind of smoothness so that $\mathcal{L}$ preserves (or improves) smoothness.

Consider now a one-sided subshift of finite type (Λ, F); i.e., with the notation used earlier, $\Lambda = \{(\xi_k)_{k\geq 0} : t_{\xi_k \xi_{k+1}} = 1 \text{ for all } k\}$ and $F(\xi_k) = (\xi_{k+1})$. We define a metric on Λ by analogy with that on Ω and take $g = \exp A$ where A is β-Hölder continuous. Then

$$(\mathcal{L}\Phi)(\xi_1, \xi_2, \dots) = \sum e^{A(\xi_0, \xi_1, \xi_2, \dots)}\Phi(\xi_0, \xi_1, \xi_2, \dots),$$

where the sum is over the ξ_0 such that $t_{\xi_0\xi_1} = 1$. Similarly

$$(\mathcal{L}^m\Phi)(\underline{\xi}) = \sum_{\underline{\eta}: F^m\underline{\eta} = \underline{\xi}} \Big[\exp \sum_{k=0}^{m-1} A(F^k\underline{\eta})\Big]\Phi(\underline{\eta}). \tag{6}$$

Now, expressions like (6) or like

$$\sum_{\underline{\eta}: F^m\underline{\eta} = \underline{\eta}} \exp \sum_{k=0}^{m-1} A(F^k\underline{\eta}) \tag{7}$$

(where the sum is over periodic points) are known in statistical mechanics as *partition functions*, and one can prove under various conditions that the logarithm of the partition function divided by m tends to a limit $P(A)$ when $m \to \infty$. Here one finds that when $\mathcal{L}$ acts on the Banach space of β-Hölder functions,

$$\lim_{m\to\infty} \|\mathcal{L}^m\|^{1/m} = e^{P(A)},$$

where $P(A)$ is defined as in (4) with τ replaced by F. Therefore $\exp P(A)$ is the spectral radius of $\mathcal{L}$. By a formula due to Nussbaum we can estimate the essential spectral radius of $\mathcal{L}$ to be

$$\leq \limsup_{m\to\infty} \|\mathcal{L}^m - E_m\|^{1/m}$$

when the E_m have finite rank. Pollicott noticed that by taking $E_m = \mathcal{L}^m P_m$ and $P_m\Phi$ a piecewise constant approximation of Φ, one gets that the essential spectral radius of $\mathcal{L}$ is $\leq \exp(-\beta + P(A))$. So the part of the spectrum which is $> \exp(-\beta + P(A))$ consists of isolated eigenvalues of finite multiplicity. In fact, because $\mathcal{L}$ is positivity preserving and F is mixing, $\exp P(A)$ is a simple eigenvalue, and there is no other eigenvalue with the absolute value $\exp P(A)$. (This is a Perron-Frobenius type result. Because of this, transfer operators are sometimes called Perron-Frobenius operators.)

Notice that (7) is something like a trace of $\mathcal{L}^m$, and because of this one can show that each eigenvalue λ of $\mathcal{L}$ contributes a factor $1/(1-\lambda z)$ to $\zeta(z)$ defined by (3). It is not obvious that the part of the spectrum $\le \exp(-\beta + P(A))$ will contribute a factor analytic for $|z| < \exp(\beta - P(A))$, but this can be proved by a trick due to Haydn (and techniques of the thermodynamic formalism). We have just outlined a modern proof of (an improved version of) Proposition 1, up to a detail: our function A depends on the one-sided sequence

$$\underline{\xi} = (\xi_0, \xi_1, \xi_2, \dots)$$

instead of on

$$\hat{\xi} = (\dots, \xi_{-1}, \xi_0, \xi_1, \dots).$$

This is, however, not a problem, because it can be proved (Livšic) that an α-Hölder function $A(\hat{\xi})$ may be rewritten as $\hat{A}(\underline{\xi}) + B(\hat{\xi}) - B(\tau\hat{\xi})$ where $\hat{A}$ is β-Hölder with $\beta = \alpha/2$, so that A and $\hat{A}$ give the same dynamical zeta function.

We have just seen how to derive analyticity properties of the zeta function associated with the weighted dynamical system $(\Lambda, F, \exp A)$ from study of the transfer operator $\mathcal{L}$ defined by (6). The same technique applies to other cases; its success depends on the choice of a Banach space $\mathcal{B}$ of "smooth" functions for which the essential spectral radius of $\mathcal{L}$ is strictly smaller than its spectral radius (i.e., $\mathcal{L}$ is quasicompact).

An important example, that of piecewise monotone maps of the interval, was treated by Baladi and Keller. Let $a = a_0 < a_1 < \dots < a_N = b$. We take Λ to be the compact set $[a, b]$ in $\mathbf{R}$ and assume that $F : \Lambda \to \Lambda$ is such that $F|(a_{i-1}, a_i)$ is continuous and strictly monotone for $i = 1, \dots, N$. Also assume that $F^m x,\ F^m y \in (a_{i(n)-1}, a_{i(n)})$ for all $m \ge 0$ implies $x = y$, and that $g : \Lambda \to \mathbf{C}$ is ≥ 0, of bounded variation with regular discontinuities. Writing

$$R = \lim_{m\to\infty} (\sup_x |\mathcal{L}^m 1(x)|)^{1/m},$$

$$\hat{R} = \lim_{m\to\infty} \Big(\sup_x \prod_{k=0}^{m-1} g(F^k x) \Big)^{1/m},$$

one obtains that $\zeta(z)$ is analytic for $|z| < R^{-1}$, meromorphic for $|z| < \hat{R}^{-1}$, and the eigenvalues λ with $\lambda > R$ of $\mathcal{L}$ acting on the functions of bounded variation correspond to poles λ^{-1} of $\zeta(z)$, with the same multiplicity. Usually, $R > \hat{R}$ and (since we assumed $g \ge 0$) R^{-1} is a pole of $\zeta(z)$.

Traces and Determinants

A trace on an algebra S over $\mathbf{C}$ is a linear operator $\mathrm{Tr} : S \to \mathbf{C}$ such that $\mathrm{Tr}\, \mathcal{M}_1 \mathcal{M}_2 = \mathrm{Tr}\ \mathcal{M}_2 \mathcal{M}_1$. In particular we shall be interested in traces on algebras generated by transfer operators (or containing them). Remember that the transfer operator $\mathcal{L}$ associated with the weighted dynamical system (M, f, g) satisfies

$$(\mathcal{L}\Phi)(x) = \sum_{y : fy = x} g(y)\Phi(y). \tag{8}$$

If $\mathcal{L}_1$, $\mathcal{L}_2$ are transfer operators associated with maps $f_1, f_2 : M \to M$ and weights $g_1, g_2 : M \to \mathbf{C}$, we have

$$(\mathcal{L}_2\mathcal{L}_1\Phi)(x) = \sum_{y : f_2 f_1 y = x} g_2(f_1 y) g_1(y) \Phi(y) \tag{9}$$

so that $\mathcal{L}_2\mathcal{L}_1$ is again a transfer operator. An example of a trace is the counting trace defined on transfer operators by

$$\mathrm{Tr}^c \mathcal{L} = \sum_{x \in \mathrm{Fix}\, f} g(x).$$

[It is readily seen from (9) that $\mathrm{Tr}^c \mathcal{L}_1\mathcal{L}_2 = \mathrm{Tr}^c \mathcal{L}_2\mathcal{L}_1$. In specific cases one would want to check that the sum in $\mathrm{Tr}^c \mathcal{L}$ converges and that $\mathrm{Tr}^c \mathcal{M}$ depends only on $\mathcal{M}$ as an operator, not on its specific representation as sum of transfer operators of the form (8).]

When we have a trace Tr we can define a determinant $\mathrm{Det}(\mathbf{1} - z\mathcal{M})$ as a formal power series

$$\mathrm{Det}(\mathbf{1} - z\mathcal{M}) = \exp\Big(- \sum_{m=1}^{\infty} \frac{z^m}{m} \mathrm{Tr}\mathcal{M}^m \Big)$$

(where $\mathbf{1}$ denotes the identity operator). If S is the algebra of $N \times N$ matrices and Tr, Det are the usual trace and determinant, the above is an identity that one can check by putting $\mathcal{M}$ in normal Jordan form. Note also that the counting determinant $\mathrm{Det}^c(\mathbf{1} - z\mathcal{L})$ constructed with the counting trace is related to the dynamical zeta function (3) by

$$\zeta(z) = 1/\mathrm{Det}^c(\mathbf{1} - z\mathcal{L}).$$

Suppose now that M is a smooth manifold and that the algebra S is generated by transfer operators of the form (8) with smooth $f : M \to M$ and $g : M \to \mathbf{C}$. We can then define a sharp trace $\mathrm{Tr}^\sharp$ such that

$$\mathrm{Tr}^\sharp \mathcal{L} = \sum_{x \in \mathrm{Fix} f} L^\sharp(x, f) g(x),$$

where $L^\sharp(x, f) = \mathrm{sgn}\det(\mathbf{1} - (T_x f)^{-1})$. [We assume here that $T_x f$ and $\mathbf{1} - T_x f$ are invertible, but we shall see later that the definition of $\mathrm{Tr}^\sharp$ extends to more general situations where $\mathrm{Fix} f$ need not be finite. Note also that if f is a diffeomorphism, $L^\sharp(x, f) = L(x, f^{-1})$.] A sharp determinant $\mathrm{Det}^\sharp(\mathbf{1} - z\mathcal{L})$ is defined correspondingly, and a sharp zeta function $\zeta^\sharp(z) = 1/\mathrm{Det}^\sharp(\mathbf{1} - z\mathcal{L})$.

Let me interrupt the discussion of traces to address an obvious problem. Following geometric intuition, we have introduced dynamical zeta functions and transfer operators associated with a weighted dynamical system (M, f, g). But the use of traces makes it natural to introduce linear combinations of transfer operators, so we lose the geometric connection with a single dynamical system. What is a natural formalism in the more general situation? Note that if there is a partition of unity (χ_ω) such that f restricted to $\mathrm{supp}\chi_\omega$ has an inverse ψ_ω, we may rewrite (8) as

$$\sum_\omega G_\omega(x)\Phi(\psi_\omega x),$$

where $G_\omega = (\chi_\omega g) \circ \psi_\omega$. We are thus led to define a *generalized transfer operator* $\mathcal{M}$ (associated with a family of weights $G_\omega : M \to \mathbf{C}$ and a family of maps $\psi_\omega : \mathrm{supp}G_\omega \to M$) by

$$(\mathcal{M}\Phi)(x) = \int d\omega\, G_\omega(x)\Phi(\psi_\omega x),$$

where $d\omega$ denotes a measure (which may be taken to be a probability measure). Linear combinations of generalized transfer operators are again generalized transfer operators: they form an algebra (under suitable conditions on the choice of the G_ω and ψ_ω). It is possible to consider $\mathcal{M}$ as a transfer operator associated with a (nonunique) random dynamical system. There is no longer a pressure associated with the generalized transfer operator $\mathcal{M}$, but writing

$$(|\mathcal{M}|\Phi)(x) = \int d\omega\, |G_\omega(x)|\Phi(\psi_\omega x),$$

we shall denote by e^P the spectral radius of $|\mathcal{M}|$ acting on bounded functions.

Let us return to the smooth situation (the G_ω and ψ_ω are C^r) and note that the sharp trace is now

$$\mathrm{Tr}^\sharp \mathcal{M} = \int d\omega \sum_{x\in \mathrm{Fix}\psi_\omega} L(x, \psi_\omega) G_\omega(x).$$

It is convenient at this point to introduce operators $\mathcal{M}^{(k)}$ acting on k-forms α such that

$$\mathcal{M}^{(k)}\alpha = \int d\omega\, G_\omega . \psi^*_\omega \alpha,$$

where $\psi^*_\omega\alpha$ is the pullback of α by ψ_ω. [If $\wedge^k(T_x\psi) : \wedge^k(T_xM) \to \wedge^k(T_{\psi x}M)$ is the extension of $T_x\psi$ to the exterior algebra of T_xM and if $(\wedge^k(T_x\psi))^* : \wedge^k(T^*_{\psi x}M) \to \wedge^k(T^*_xM)$ denotes its transpose, we write $(\psi^*\alpha)(x) = (\wedge^k(T_x\psi))^*\alpha(\psi x)$.] In particular $\mathcal{M}^{(0)}$ reduces to $\mathcal{M}$. Following Atiyah and Singer, we define now a flat trace $\mathrm{tr}^\flat_k$ such that

$$\mathrm{Tr}^\flat_k \mathcal{M}^{(k)} = \int d\omega \sum_{x\in\mathrm{Fix}\psi_\omega} \frac{G_\omega(x)\mathrm{tr}_k(\wedge^k(T_x\psi_\omega))}{|\det(\mathbf{1} - T_x\psi_\omega)|},$$

where tr_k and det are the finite-dimensional trace and determinant. [Writing $\mathcal{M}^{(k)}$ as the limit of a regularized operator with kernel $M(x, y)$, we obtain $\mathrm{Tr}^\flat_k \mathcal{M}^{(k)}$ as the limit of $\int dx\, \mathrm{tr}_k M(x, x)$.] It is readily seen that

$$\mathrm{Tr}^\sharp \mathcal{M} = \sum_{k=0}^n (-1)^k \mathrm{Tr}^\flat_k \mathcal{M}^{(k)}$$

so that

$$\mathrm{Det}^\sharp(\mathbf{1} - z\mathcal{M}) = \prod_{k=0}^n \mathrm{Det}^\flat_k(\mathbf{1} - z\mathcal{M}^{(k)})^{(-1)^{k+1}}.$$

3. *Proposition.*[*] *Let M be a compact Riemann manifold. We assume that $G_\omega : M \to \mathbf{C}$ and $\psi_\omega : X_\omega \to M$ (where X_ω is a δ-neighborhood of* supp G_ω*) are* C^r, $r \geq 1$, *depending measurably on ω, and that*

$$\int d\omega\, \|G_\omega\|_r < \infty, \quad \sup_\omega \|\psi_\omega\|_r < \infty.$$

Also assume that there is $\theta \in (0, 1)$ such that

$$\mathrm{dist}(\psi_\omega x, \psi_\omega y) \leq \theta\, \mathrm{dist}(x, y)$$

for all x, y, ω. Then the part of the spectrum of $\mathcal{M}$ in $\{\lambda : |\lambda| > \theta^r e^P\}$ consists of isolated eigenvalues of finite multiplicities. Furthermore, $\mathrm{Det}^\flat_0(\mathbf{1} - z\mathcal{M})$ *converges in $\{z : |z|\theta^r e^P < 1\}$, and its zeros there are precisely the inverses of the eigenvalues of $\mathcal{M}$ with the same multiplicity.*

There are results similar to the above proposition for $\mathcal{M}^{(k)}$ and $\mathrm{Det}^\flat_k(\mathbf{1} - z\mathcal{M}^{(k)})$. It follows in particular that $1/\mathrm{Det}^\sharp(\mathbf{1} - z\mathcal{M})$ is meromorphic for $|z| < \theta^{-r}e^{-P}$. Note that for contracting ψ we have $L(x, \psi) = 1$; hence $\mathrm{Det}^\sharp = \mathrm{Det}^c$, and we obtain results for the dynamical zeta functions $\zeta(z)$ associated with smooth expanding maps (first studied by Tangerman). Proposition 3 also applies to a rational map F if it is hyperbolic, i.e., uniformly expanding in a neighborhood of the Julia set J (the closure of the set of repelling periodic orbits).

Proposition 3 is a nonstandard extension of the theory of Fredholm determinants. In its simplest form, Fredholm's theory applies to complex continuous kernels $K(x, y)$, where x and y vary over a bounded interval $[a, b]$. The formula

$$K\phi(x) = \int_a^b K(x, y)\phi(y)dy$$

defines a compact operator on the Banach space $\mathcal{B}$ of complex continuous functions on $[a, b]$ with the sup norm. The operators K as above form an algebra, with a trace

[*] *This result is proved in* [4]. *The condition* $\sup_\omega \|\psi_\omega\|_r < \infty$ *is missing in* [4], *but some form of this condition is used in Remark 3.3 of that paper. Note that we may take fractional $r = s + \alpha$, meaning that the s-th derivative is α-Hölder.*

$$\mathrm{Tr}_F K = \int_a^b K(x,x)dx,$$

and one can define a *Fredholm determinant* by

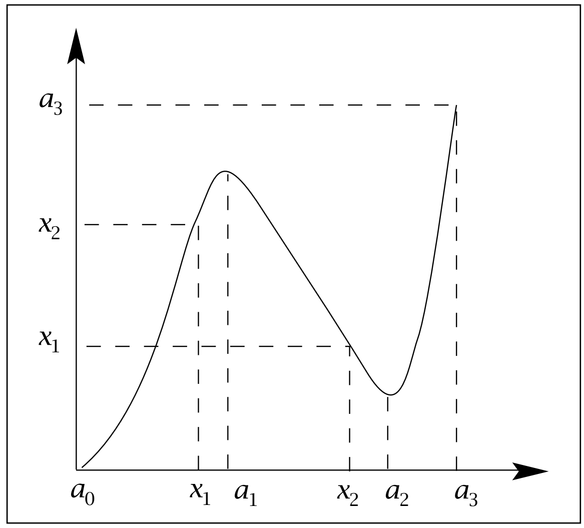
$$\mathrm{Det}_F(\mathbf{1} - zK) = \exp\left(-\sum_{m=1}^{\infty} \frac{z^m}{m} \mathrm{Tr}_F K^m\right)$$

or some equivalent formula. This determinant is an entire function of z which has a zero at λ^{-1} precisely when λ is an eigenvalue of K (the order of the zero and of the eigenvalue are the same). Fredholm's theory has been put on a more conceptual basis by Grothendieck, using kernels in the topological tensor product $\mathcal{B}^* \hat{\otimes} \mathcal{B}$ of a Banach space $\mathcal{B}$ and its dual $\mathcal{B}^*$. Grothendieck's extension of Fredholm's theory applies in particular to holomorphy-improving operators (these send a function holomorphic in D to a function holomorphic in D', where D is relatively compact in D'). The Fredholm-Grothendieck determinant $\mathrm{Det}_F(\mathbf{1} - zK)$ is an entire function of z, but note that in Proposition 3, $\mathrm{Det}_0^\flat(\mathbf{1} - z\mathcal{M})$ has in general a finite radius of convergence and that $\mathcal{M}$ is not a compact operator.

Proposition 3 applies to expanding maps. What about hyperbolic maps (say Anosov) on a compact manifold M? For such maps there is an invariant family of submanifolds of M called stable manifolds, which are uniformly contracted by the map. These manifolds are smooth, but the stable manifold through x does not depend smoothly on x, only Hölder continuously. For this reason one cannot readily extend to the smooth situation what was done (see Proposition 2 above) in the Hölder setting. [The case of C^ω Anosov maps in two dimensions has been elegantly treated by Rugh. The general case has been discussed by Kitaev, but his paper is difficult. Work by Fried on the subject remains unpublished. There is also recent work of Blank and of Keller and Liverani on transfer operators in two dimensions.] When the stable manifolds form a smooth family, an extension of Proposition 3 to the hyperbolic situation works well. This happens in particular for the geodesic flow on a manifold of constant negative curvature, where everything is C^ω: the zeta functions are quotients of Grothendieck-type determinants, and thus meromorphic in $\mathbf{C}$ (Ruelle). This agrees with what is known about Selberg zeta functions and extends to other situations (Mayer, Patterson).

Kneading Determinants

Milnor and Thurston have studied continuous piecewise monotone maps of the interval $[a_0, a_N]$ to itself that are strictly monotone on subintervals $[a_{i-1}, a_i]$, where $a_0 < a_1 < \ldots < a_N$. A (slightly modified) zeta function ζ_{MT} which counts periodic orbits with a weight 1 satisfies



A continuous piecewise monotone map of the interval $[a_0, a_3]$ to itself, with three subintervals where the function is strictly monotone. $\{x_1, x_2\}$ is a periodic orbit of order 2.

$$\zeta_{MT}(z).\Delta(z) = 1\,, \tag{10}$$

where Δ is the determinant of a certain $(N-1)\times(N-1)$ matrix called the *kneading matrix*. The elements of the kneading matrix are power series in z with coefficients $0, \pm 1$ determined in terms of the signs of the $f^m a_i - a_j$. In particular, $\zeta_{MT}(z)$ is meromorphic in the unit disc. Can one extend the combinatorial identity (10) to dynamical zeta functions with weights? Baladi and I obtained an extension where Δ is replaced by a functional determinant.

We consider generalized transfer operators $\mathcal{M}$ acting on the Banach space $\mathcal{B}$ of functions of bounded variation on $\mathbf{R}$, so that

$$(\mathcal{M}\Phi)(x) = \sum_\omega G_\omega(x)\Phi(\psi_\omega x).$$

Here the $G_\omega : \mathbf{R} \to \mathbf{C}$ are of bounded variation, compactly supported, and (for simplicity) continuous; ψ_ω is a homeomorphism of an interval of $\mathbf{R}$ containing supp G_ω to an interval of $\mathbf{R}$, and we assume $\sum_\omega \mathrm{Var}\, G_\omega < \infty$. Write $\epsilon_\omega = +1$ ($\epsilon_\omega = -1$) if ψ_ω is increasing (decreasing). The operators $\mathcal{M}$ form an algebra $\mathcal{A}$ with an involution $\mathcal{M} \to \widehat{\mathcal{M}}$ where

$$(\widehat{\mathcal{M}}\Phi)(x) = \sum_\omega \epsilon_\omega G_\omega(\psi_\omega^{-1}x)\Phi(\psi_\omega^{-1}x)$$

and, using the sup norm $\|\cdot\|_0$, we write

$$R = \lim_{m\to\infty}(\|\mathcal{M}^m\|_0)^{1/m}, \quad \hat{R} = \lim_{m\to\infty}(\|\widehat{\mathcal{M}}^m\|_0)^{1/m}.$$

It turns out that, for the spectrum of $\mathcal{M}$ acting on $\mathcal{B}$,

$$\hat{R} \leq \text{spectral radius of } \mathcal{M} \leq \max(R, \hat{R})$$
$$\text{essential spectral radius of } \mathcal{M} \leq \hat{R}.$$

The interesting case is when $R \neq \hat{R}$. In particular, if $\hat{R} < R$ and all G_ω are ≥ 0, then R is an eigenvalue of $\mathcal{M}$.

The sharp trace $\mathrm{Tr}^\#$ defined earlier can be extended to $\mathcal{A}$ by writing

$$\mathrm{Tr}^\# \mathcal{M} = \sum_\omega \int \frac{1}{2} \mathrm{sgn}(\psi_\omega(x) - x)\, dG_\omega(x),$$

where $\mathrm{sgn}\ \xi = \xi/|\xi|$ if $\xi \neq 0$ and $\mathrm{sgn}\ 0 = 0$. We can then define the zeta functions

$$\zeta(z) = \frac{1}{\mathrm{Det}^\#(\mathbf{1} - z\mathcal{M})}, \quad \hat{\zeta}(z) = \frac{1}{\mathrm{Det}^\#(\mathbf{1} - z\widehat{\mathcal{M}})},$$

and interestingly we have the *functional equation* $\zeta(z)\hat{\zeta}(z) = 1$ (from $\mathrm{Tr}^\# \mathcal{M} + \mathrm{Tr}^\# \widehat{\mathcal{M}} = 0$).

A bounded nonatomic measure on $\mathbf{R}$ is given by

$$\mu(dx) = \sum_\omega |dG_\omega(x)| + \sum_\omega |dG_\omega(\psi_\omega^{-1} x)|.$$

We define now a *kneading operator* $\mathcal{D}$ on $L^2(\mu)$ by

$$(\mathcal{D}\phi)(y) = \sum_\omega \int \phi(x)\, d(zG_\omega(x)) \times \{[(\mathbf{1} - z\mathcal{M})^{-1} \frac{1}{2} \mathrm{sgn}(\cdot - y)](\psi_\omega x)\}$$

and similarly an operator $\widehat{\mathcal{D}}$. The kernel of $\mathcal{D}$ is given by

$$\mathcal{D}_{xy} = \sum_{k=1}^\infty z^k \sum_{\omega_1, \cdots, \omega_k} \frac{dg_{\omega_1}(x)}{\mu(dx)} \cdot g_{\omega_2}(\psi_{\omega_1} x) \dots g_{\omega_k}(\psi_{\omega_{k+1}} \dots \psi_{\omega_1} x) \times \frac{1}{2} \mathrm{sgn}(\psi_{\omega_k} \dots \psi_{\omega_1} x - y).$$

It turns out that $\mathcal{D}$ is Hilbert-Schmidt, and one can define a functional determinant

$$\mathrm{Det}(\mathbf{1} + \mathcal{D}) = \exp\Big(\int \mu(dx)\, \mathcal{D}_{xx} + \sum_{m=2}^\infty \frac{(-1)^{m-1}}{m} \mathrm{Tr}\mathcal{D}^m\Big).$$

What corresponds to the Milnor-Thurston determinant is here $\mathrm{Det}(\mathbf{1} + \widehat{\mathcal{D}})$; i.e., one can prove the identity

$$\zeta(z) = \mathrm{Det}(\mathbf{1} + \widehat{\mathcal{D}})^{-1}.$$

From this one can deduce that the determinant $\mathrm{Det}^\#(\mathbf{1} - z\mathcal{M}) = \mathrm{Det}(\mathbf{1} + \widehat{\mathcal{D}})$ is holomorphic for $|z| < \hat{R}^{-1}$ and that its zeros there are the λ^{-1} where $\hat{R} < |\lambda| < R$ and λ is an eigenvalue of $\mathcal{M}$ (of the same multiplicity).

Extensions of the theory of kneading determinants to dimension greater than 1 have been studied (Baladi, Kitaev, Ruelle, Semmes, Baillif) and are currently an active area of research, but only partial results have been obtained so far.

Some Loose Ends

Counting periodic orbits with weights is a natural idea. And we have seen that it relates to very different areas of mathematics: thermodynamic formalism, hyperbolic dynamics, Selberg zeta functions, Grothendieck-Fredholm determinants, kneading determinants, etc. The "hyperbolic" part of the theory of dynamical zeta functions is excellently presented in the monograph of Parry and Pollicott [2], which gives more details on the relation with the thermodynamic formalism than could be given here. For further developments we refer to Baladi's monograph [1], which discusses in particular the relation between spectral properties of transfer operators and the decay of correlations. A discussion of the decay of correlations would have taken us too far afield, but this is an important topic which has progressed in recent years, thanks to the work of Dolgopyat on hyperbolic flows and the very general ergodic results of Young [5]. Using the extensive bibliography of [1], the interested reader can get access to many other questions: for instance, the surprising results of Mayer on the continued fraction transformation and the modular surface, and the very explicit formulas obtained by Levin, Sodin, and Yuditskii in the study of Julia sets.

Dynamical zeta functions and the related concepts discussed in this article form a rather open field of investigation. Some astonishing developments have occurred in the past. And new technical or structural ideas might again drastically change our view of the subject in the future.

Acknowledgments

I am indebted to Viviane Baladi for critical reading of the present article and also to the referee and editor for a number of helpful suggestions. The article was written in part at Rutgers, in part at the IHÉS, and in part at the IMPA (Rio).

References

[1] V. Baladi, *Positive transfer operators and decay of correlations*, World Scientific, Singapore, 2000.

[2] W. Parry and M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque **187–188**, Soc. Math. de France, Paris, 1990.

[3] D. Ruelle, *Thermodynamic formalism*, Addison-Wesley, Reading, MA, 1978.

[4] ____, An extension of the theory of Fredholm determinants, *Publ. Math. IHÉS* **72** (1990), 175–193.

[5] L.-S. Young, Developments in chaotic dynamics, *Notices Amer. Math. Soc.* **45** (1998), 1318–1328.