Recycling of strange sets: I. Cycle expansions

Roberto Artuso†§, Erik Aurell‡ and Predrag Cvitanović†∥

† Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark
‡ Institute of Theoretical Physics, Chalmers University of Technology and University of Göteborg, S-412 96 Göteborg, Sweden

Abstract. The strange sets which arise in deterministic low-dimensional dynamical systems are analysed in terms of (unstable) cycles and their eigenvalues. The general formalism of cycle expansions is introduced and its convergence discussed.

1. Introduction

The goal of the present series of papers is the development of a perturbation theory of the low-dimensional deterministic chaos of predictive quality comparable to that of the traditional perturbation expansions for nearly integrable systems. In the traditional approach the integrable motions are used as zeroth-order approximations to physical systems, and weak nonlinearities are then accounted for perturbatively. For strongly nonlinear, non-integrable systems such expansions fail completely; the asymptotic time phase space exhibits amazingly rich structure which is not at all apparent in the integrable approximations. However, hidden in this apparent chaos is a rigid skeleton, a tree of *cycles* (periodic orbits) of increasing lengths and self-similar structure. The important role played by periodic orbits was already noted by Poincaré [1], and has been at the core of much of the mathematical work on the theory of the dynamical systems [2] ever since. The insight of the modern dynamical systems theory [3] is that the zeroth-order approximations to the harshly chaotic dynamics should be very different from those for nearly integrable systems: a good starting approximation here is the linear stretching and folding of a baker's map, rather than the winding of a harmonic oscillator.

The present work is a physics application of the dynamical systems theory developed in [3–7]; we refer the reader to the above literature for a survey of rigorous results. Computations with such systems require techniques reminiscent of statistical mechanics; however, no probabilities are introduced and the actual calculations are crisply deterministic. The perturbation theory developed here is based on the observation that the motion in dynamical systems of a few degrees of freedom is often organised around a few *fundamental* cycles. The strategy will be to express averages over chaotic

§ Present address: Dipartimento di Fisica dell'Università and INFN, Via Celoria 16, I-20133 Milano, Italy.
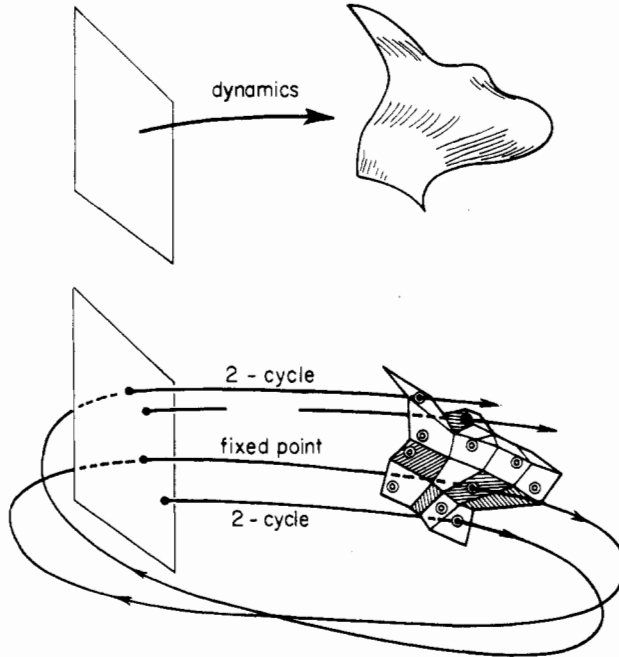∥ Carlsberg Fellow.

Figure 1. Tessellation of a dynamical system by cycles; a smooth flow is approximated by a piecewise linear mapping, with each 'face' centred on a periodic point.

phase space regions in terms of short unstable periodic orbits, with the small expansion parameter being the non-uniformity of the flow (here referred to as *curvature*) across neighbourhoods of periodic points. The emphasis will be on the practical applications of cycle expansions, at some expense to the mathematical rigour. We are interested in the convergence of cycle expansions in generic settings, i.e. situations in which neither the symbolic dynamics is strictly controlled (there is no finite Markov partition), nor is the system uniformly hyperbolic (the phase space is a mixture of stability islands and chaotic regions). Our results will not depend on assumptions about the existence of invariant measures or structural stability of strange sets.

We shall refer to the closure of the union of periodic points as the *strange set*; all calculations undertaken here are carried out on strange sets. Our main computational tool will be the *cycle expansions* [8] of the dynamical ζ functions [6]:

$$1/\zeta = \prod_p (1 - t_p) = 1 - \sum_f t_f - \sum_n c_n. \tag{1}$$

The *fundamental* cycles t_f have no shorter approximants; they are the 'building blocks' of the dynamics in the sense that all longer orbits can be approximately pieced together from them. They code exactly the topology of the strange set, and serve as the starting approximation to its scalings. *A priori* it is far from obvious that a few finite cycles suffice to describe the infinity of orbits characteristic of a chaotic dynamical system. We will show here how this infinity of orbits can be resummed and re-expressed in a form in which the short fundamental cycles dominate, and the errors arising from neglect of longer cycles can be controlled with exponential (and occasionally rather impressive) accuracy.

Periodic points are skeletal in the sense that even though they are determined at finite time, they remain there forever. One can visualise the description of a chaotic dynamical system in terms of cycles as a tessellation of the dynamical system (figure 1) with smooth flow approximated by the skeleton of periodic points of period length n, each region V_i centred on a periodic point x_i, and the size of the region determined by the linearisation of the flow around the periodic point. Instead of temporal averages (a long trajectory which explores the phase space ergodically) we shall always work with finite time, topologically partitioned space averages. The periodic points are dense on the asymptotic strange set, and their number increases exponentially with the cycle length. As we shall see, this exponential proliferation of cycles is not as daunting as one might fear; as a matter of fact, all our computations are carried out in the $n \to \infty$ limit. The infinity of cycles required by the exact dynamics will be approximated by shadowing long orbits with sequences of nearby periodic orbits of finite lengths. Orbits that follow the same symbolic dynamics, such as orbit $\{ab\}$ and a 'pseudo-orbit' $\{a\}\{b\}$, lie physically close; longer and longer orbits resolve the dynamics with finer and finer resolution in the phase space. If the weights t_a associated with the orbits are multiplicative along the flow (for example, products of derivatives) and the flow is smooth, the combination $t_a t_b / t_{ab} - 1$ falls off *exponentially* with the cycle length. The *curvature* corrections c_n in (1) are built from such combinations, and the cycle expansions are therefore highly convergent. We show that this is the case even for non-hyperbolic dynamical systems (systems with orbits of marginal stability), provided that the averaging is done in the 'hyperbolic phase'.

The paper is organised as follows: in section 2 we derive a ζ function formula for a physically measurable quantity, the escape rate from a repeller, and in section 3 we repeat the derivation in a transfer operator formalism. In section 4 we introduce our main tool, the cycle expansions. In section 5 we discuss the form the cycle expansions take for pruned symbolic dynamics; in section 6 we apply them to calculation of topological entropies, and in section 7 we discuss their convergence. In section 8 we introduce the notion of the stability of a strange set, and in section 9 we explain how the the standard thermodynamic averages can be computed in terms of cycle expansions.

The present paper concentrates on the general properties of the cycle expansions; in the following paper (hereafter referred to as II) we apply the cycle expansions to a series of examples of low-dimensional chaos: 1D strange attractors, the period-doubling repeller, the Hénon-type maps and the mode locking intervals for circle maps. Beyond the examples discussed in II, the cycle expansions have also been applied to the irrational windings set of the critical circle maps [9], to the Hamiltonian period-doubling repeller [10], to a Hamiltonian three-disk pinball [11], to the three-disk quantum scattering resonances [12, 13] and to the extraction of correlation exponents [14]. Feasibility of analysis of experimental strange sets in terms of cycles is discussed in [8].

2. Escape rates

A repeller escape rate is an eminently measurable quantity. The experimental measurement consists in shooting many projectiles into a non-confining potential and estimating the asymptotic escape rate; the task of the theory is to predict this rate. We shall show here that such escape rates (and other chaotic averages) can be predicted to very

high accuracy with rather little computation. We start with a simple one-dimensional repeller example, comment on its generalisation to continuous d-dimensional flows and then repeat the derivation in a more general transfer-operator setting in the next section.
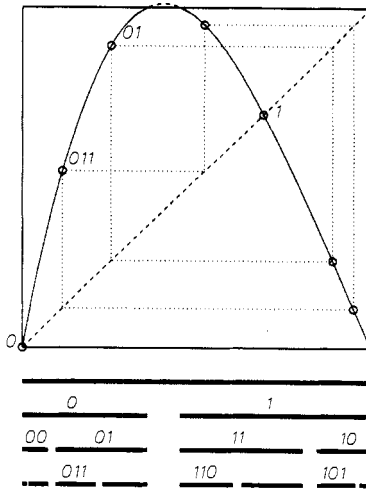

Figure 2. A hierarchical covering of a strange repeller by intervals that survive one, two and three iterations of a unimodal repelling map $f(x)$. Indicated are the binary itineraries of the neighbourhoods, the fixed points $\bar{0}$ and $\bar{1}$ and the $\overline{011}$ cycle. A 'daughter' interval l_d expands into the 'mother' interval at rate well approximated by $f'(x_d)$: for example $l_{01}/l_{101} \approx f'(x_{101})$.

Take the unimodal repeller of figure 2, with $f(x_c) > x_{max}$, and sprinkle the unit interval with a smooth distribution of starting values of x. In the first iteration an interval around the maximum x_c escapes, in the second iteration its two pre-images escape, and so on. At time n the survivors are divided into 2^n distinct neighbourhoods: the ith neighbourhood consists of all points x which follow the itinerary $i = \epsilon_1\epsilon_2\epsilon_3\ldots\epsilon_n$, with $\epsilon_k = 0$ if $f^{(k)}(x) < x_c$, and $\epsilon_k = 1$ if $f^{(k)}(x) > x_c$. Let l_i be the width of such a neighbourhood, or, more generally, the fraction of inital x placed into the ith neighbourhood. The fraction of the initial x which survive n iterations is given by

$$\Gamma_n = \sum_i^{(n)} l_i. \tag{2}$$

The map is smooth, and its derivative bounded and everywhere expanding, $1 < |\Lambda_{min}| \le |df/dx| \le |\Lambda_{max}|$, so each interval in (2) is bounded by $|\Lambda_{max}^{-n}| \le l_i \le |\Lambda_{min}^{-n}|$. Replacing l_i in (2) by its over (under) estimates in terms of $|\Lambda_{max}|$, $|\Lambda_{min}|$ immediately leads to exponential bounds $(2/|\Lambda_{max}|)^n \le \Gamma_i \le (2/|\Lambda_{min}|)^n$. A finer graining and counting of scales would lead to improved bounds—establishing these bounds is in a sense precisely the goal of the present series of papers. Hence one expects the sum (2) to fall off exponentially with n, and tend to a limit

$$\Gamma_n = e^{-n\gamma_n} \rightarrow e^{-n\gamma}. \tag{3}$$

$\gamma = 1/T$ is the *escape rate* from the repeller; T is the *asymptotic lifetime* of a random initial x. We shall now show that this asymptotic escape rate can be extracted from a

highly convergent *exact* expansion by reformulating the sum (2) in terms of unstable periodic orbits.

Each neighbourhood i in figure 2 contains a periodic point x_i. The finer the intervals, the smaller the variation in slope across them, and the expansion of l_i onto the unit interval in n iterations is well approximated by the stability of the periodic point x_i, $l_i = a_i/|\Lambda_i|$. Here

$$\Lambda_i = \frac{d}{dx} f^{(n)}(x_i) = \prod_{k=0}^{n-1} f'(f^{(k)}(x_i)) \tag{4}$$

is the derivative evaluated along the periodic orbit, and a_i is a prefactor defined by

$$a_i = l_i|\Lambda_i|. \tag{5}$$

To proceed with the derivation of the ζ function we need the *hyperbolicity* assumption: for large n the prefactors $a_i \approx O(1)$ are overwhelmed by the exponential growth of Λ_i, so we neglect them. The a_i reflect a particular distribution of starting values of x; the asymptotic trajectories are strongly mixed by bouncing chaotically around the repeller and we expect them to be insensitive to smooth variations in the initial distribution. If the hyperbolicity assumption is justified, we can replace l_i in (2) by $1/\Lambda_i$ and form a formal sum over all periodic orbits of all lengths:

$$\Omega(z) = \sum_{n=1}^{\infty} z^n \sum_{i}^{(n)} |\Lambda_i|^{-1} = z/|\Lambda_0| + z/|\Lambda_1| + z^2/|\Lambda_{00}|$$
$$+ z^2/|\Lambda_{01}| + z^2/|\Lambda_{10}| + z^2/|\Lambda_{11}| + z^3/|\Lambda_{000}| + z^3/|\Lambda_{001}| + \dots . \tag{6}$$

For sufficiently small z this sum is convergent. As for large n the nth-level sum (2) tends to the limit $e^{-n\gamma}$, the escape rate γ is determined by the smallest $z = e^{\gamma}$ for which (6) diverges:

$$\Omega(z) \approx \sum_{n=1}^{\infty} (z\,e^{-\gamma})^n = \frac{z\,e^{-\gamma}}{1 - z\,e^{-\gamma}}. \tag{7}$$

This observation motivates the introduction of the sum (6). Rather than attempting to extrapolate the escape rate from the finite n sums (2), we shall determine γ from the singularities of (6).

If a trajectory retraces itself r times, its derivative is Λ_p^r, where p is a *prime* cycle. A prime cycle is a single traversal of the orbit; its label is a non-repeating symbol string. There is only one prime cycle for each cyclic permutation class. For example, $p = \overline{0011} = \overline{1001} = \overline{1100} = \overline{0110}$ is prime, but $\overline{0101} = \overline{01}$ is not. (a bar over a finite block of symbols denotes a symbol sequence with infinitely repeating basic block). The stability of a cycle is (by the chain rule, see (4)) the same everywhere along the orbit, so each prime cycle of length n_p contributes n_p terms to the sum (6). Hence (6) can be rewritten as

$$\Omega(z) = \sum_{p} n_p \sum_{r=1}^{\infty} (z^{n_p}|\Lambda_p^{-1}|)^r = \sum_{p} \frac{n_p z^{n_p}|\Lambda_p^{-1}|}{1 - z^{n_p}|\Lambda_p^{-1}|}$$

where the index p runs through all distinct *prime* cycles. The $n_p z^{n_p}$ factors in the sum suggest rewriting it as a derivative $\Omega(z) = -z(d/dz) \sum \ln(1 - z^{n_p}/|\Lambda_p|)$. Hence $\Omega(z)$ is a logarithmic derivative of the infinite product

$$1/\zeta(z) = \prod_p (1 - z^{n_p}|\Lambda_p^{-1}|). \tag{8}$$

This is an example of a dynamical ζ function [6]. The name is motivated by the (purely formal) similarity of the infinite product to the Euler product representation of the Riemann ζ function.

The above ζ function can be immediately generalised to higher dimensions by defining [15] the escape rate from a finite enclosure V around a d-dimensional repeller by

$$e^{-n\gamma_n} = \frac{\int_V dx \, dy \, \delta(y - f^{(n)}(x))}{\int_V dx}. \tag{9}$$

An argument similar to the one that leads to the neglect of the a_i prefactors in (5) leads to a replacement of (9) by a periodic orbit sum

$$e^{-n\gamma_n} = \int_V dx \, \delta(x - f^{(n)}(x)) = \sum_i^{(n)} \frac{1}{|\det(1 - \mathbf{J}^{(n)}(x_i))|} = \sum_i^{(n)} \frac{1}{|\prod_{a=1}^d (1 - \Lambda_i^a)|} \tag{10}$$

where

$$\mathbf{J}^{(n)}(x_i) = \prod_{j=0}^{n-1} \mathbf{J}(f^{(j)}(x_i)) \qquad J_{kl}(x) = \frac{\partial}{\partial x_l} f_k(x) \tag{11}$$

is the i-cycle $[d \times d]$ Jacobian matrix, and $\Lambda_i^1, \Lambda_i^2, \ldots, \Lambda_i^d$ are its eigenvalues. If $|\Lambda_i^a| \neq 1$ (no eigenvalues are marginal), in the large-n limit only the expanding eigenvalues contribute, and (10) becomes

$$e^{-n\gamma_n} = \sum_i^{(n)} \frac{1}{|\Lambda_i|} \tag{12}$$

where $\Lambda_i = \prod_a^{\exp} \Lambda_i^a$ is the product of the expanding eigenvalues. Hence the ζ function (8) is correct for d-dimensional maps as well, with Λ_p interpreted as the product of the expanding eigenvalues.

The ζ function for continuous flows (which we shall not need for the applications considered here and in paper II) is given in [11, 16].

Expression (8) is the main result of this section; the problem of estimating the asymptotic escape rates from finite n sums such as (2) is now reduced to studying the singularities of the ζ function (8). The escape rate is related by (7) to a divergence of $\Omega(z)$, and $\Omega(z)$ diverges whenever $1/\zeta(z)$ or $\zeta(z)$ has a zero.

We conclude this section by a general comment on the relation of the finite sum (2) to the dynamical ζ function (8). Not so long ago most physicists were inclined to believe that given a deterministic rule, a sum like (2) could be evaluated to any desired precision. For short finite times this is indeed true: every interval in (2) can

be accurately determined, and there is no need for an elaborate theory. However, if a dynamical system is unstable, local variations in initial conditions grow exponentially and in finite time attain the size of the system. The difficulty with estimating the $n \to \infty$ limit from (2) is then at least twofold:

(1) due to the exponential growth in number of intervals, and the exponential decrease in attainable accuracy, the maximal n attainable experimentally or numerically is in practice of order of something between 5 and 20;

(2) the pre-asymptotic sequence γ_n is not unique, because Γ_n are not scale invariant, and because in general the intervals l_i in the sum (2) should be weighted by the probability distribution of initial x_0. For example, a rescaling $l_i \to \alpha l_i$ introduces $1/n$ corrections in γ_n defined by the sum (2): $\gamma_n \to \gamma_n - \ln \alpha/n$. This is usually fixed by extracting γ_n from successive ratios $e^{\gamma_n} \equiv \Gamma_n/\Gamma_{n+1}$. In contrast, the ζ function (8) is already invariant under *all* smooth nonlinear conjugacies $x \to h(x)$, not only linear rescalings, and requires no $n \to \infty$ extrapolations.

The pleasant surprise implicit in (8) is that the infinite time behaviour of an unstable system will be as easy to determine as the short time behaviour. The only critical step in the derivation of the ζ function was the hyperbolicity assumption, i.e. assumption of exponential growth for all parts of the strange set. By dropping the prefactors (5), we have given up on any possibility of recovering the precise distribution of starting x (which should anyhow be impossible due to the exponential growth of errors), but in return gained an effective description of the asymptotic behaviour of the system.

3. Transfer operators

The formalism of the preceding section is vastly more powerful than the 1D repeller example might suggest—the technique is meant to apply to *any* average in which the weight assigned to a trajectory is *multiplicative* along the dynamical flow. In order to place the method in this larger setting, we now rederive the dynamical ζ function (8) by the *transfer operator* technique.

Consider the classical example of a fractal [17], the Cantor set. The set is generated by a single rule: replace a mother interval l by two daughters of length $l/3$; repeat this replacement *ad infinitum*. Given the rule, one can immediately compute the Hausdorff dimension; at the nth level the set can be covered with 2^n intervals of size 3^{-n}, hence $D = \log 2/\log 3$. A transfer operator is a generalisation of such a rule to strange sets for which the dynamics generates an infinity of scales, not just a single scale as in the Cantor set case. For example, for a repeller like the one illustrated in figure 2 the dynamics associates with each 'mother' interval l_m, $m = \epsilon_2 \epsilon_3 \ldots \epsilon_n$, two 'daughter' intervals l_d, $d = 0\epsilon_2 \ldots \epsilon_n$, $1\epsilon_2 \ldots \epsilon_n$, at the next level of resolution. The transfer operator appropriate to the evaluation of (2) is defined by the set of daughter/mother ratios

$$T_{dm} = l_d/l_m. \tag{13}$$

For the Cantor set $T_{dm} = 1/3$ for all d; for a generic dynamical strange set T_{dm} takes on an infinity of values. The sum (2) can now be expressed in terms of products of transfer operators:

$$\Gamma_n = \sum_{\epsilon_1 \epsilon_2 \ldots \epsilon_n} T_{\epsilon_1 \epsilon_2 \ldots \epsilon_n, \epsilon_2 \ldots \epsilon_n} T_{\epsilon_2 \ldots \epsilon_n, \epsilon_3 \ldots \epsilon_n} \cdots T_{\epsilon_n,}. \tag{14}$$

As it stands, this is a purely formal rewrite of (2); the 'mother' to 'daughters' relations place the pieces of a strange set onto a hierarchical tree, and that can be done in various ways. To proceed, we require that the tree provide a hierarchical nesting of the scaling ratios in the following sense: the value of $T_{\epsilon_1\epsilon_2...\epsilon_n,\epsilon_2...\epsilon_n}$ should depend strongly on the head of the symbol sequence $\epsilon_1\epsilon_2...$, and weakly on the tail $...\epsilon_{n-1}\epsilon_n$. More precisely, we assume that the specification of first k symbols determines $T_{\epsilon_1\epsilon_2...\epsilon_n,\epsilon_2...\epsilon_n}$, $n > k$, within accuracy Δ_k

$$T_{\epsilon_1\epsilon_2...\epsilon_n,\epsilon_2...\epsilon_n} = \tilde{T}^{(k)}_{\epsilon_1\epsilon_2...\epsilon_k,\epsilon_2...\epsilon_{k+1}} + O(\Delta_k) \tag{15}$$

and that $|\Delta_k|$ decrease monotonically towards zero with increasing k. Here $\tilde{T}^{(k)}$ is an approximate 'mean' scaling for all T_{dm} with the same first k symbols. Replacing the infinite number of scaling ratios (13) by a finite matrix $\tilde{T}^{(k)}$ amounts to approximating the strange set by a Cantor set with a finite number of scales.

An example of such hierarchy is the repeller of the preceding section, for which $T_{dm} \approx 1/|f'(x)|$, where $f'(x)$ is a slope of the mapping evaluated at a point x inside the dth neighbourhood. With the labelling conventions of figure 2, the points whose itineraries have the same head $\epsilon_1\epsilon_2...\epsilon_n$ are spatially close, and hence the associated derivatives and transfer matrix elements are close.

Now we can study the transfer operator T as a limit of $\tilde{T}^{(k)}$ finite matrix approximations. For example, for the binary labelled repeller of figure 2, $k = 2$ level approximation to T is given by

$$\tilde{T}^{(2)} = \begin{bmatrix} \tilde{T}_{00,00} & \tilde{T}_{00,01} & 0 & 0 \\ 0 & 0 & \tilde{T}_{01,10} & \tilde{T}_{01,11} \\ \tilde{T}_{10,00} & \tilde{T}_{10,01} & 0 & 0 \\ 0 & 0 & \tilde{T}_{11,10} & \tilde{T}_{11,11} \end{bmatrix}. \tag{16}$$

$\tilde{T}$ is in general a sparse matrix, as the only non-vanishing entries in the $m = \epsilon_2\epsilon_3...\epsilon_{k+1}$ column of $\tilde{T}_{dm}$ are in the rows $0\epsilon_2...\epsilon_k$ and $1\epsilon_2...\epsilon_k$.

In the kth-order approximation the sum (14) is given by

$$\Gamma_n \approx \Gamma_n^{(k)} = \sum_{\epsilon_1\epsilon_2...\epsilon_n} \tilde{T}^{(k)}_{\epsilon_1\epsilon_2...\epsilon_k,\epsilon_2...\epsilon_{k+1}} \tilde{T}^{(k)}_{\epsilon_2...\epsilon_{k+1},\epsilon_3...\epsilon_{k+2}} \cdots T_{\epsilon_n,}$$

$$= \sum_{\epsilon_1\epsilon_2...\epsilon_k} \sum_{\delta_1...\delta_k} (\tilde{\mathbf{T}}^{n-k}_{(k)})_{\epsilon_1\epsilon_2...\epsilon_k,\delta_1...\delta_k} l_{\delta_1...\delta_k}. \tag{17}$$

Here l is the vector of all intervals l_i at the kth level. It plays the same role as the prefactors a_i in (5) of the preceding section; in the $n \gg k$ limit, the kth-level approximation (17) is dominated by the leading eigenvalue of $\tilde{T}^{(k)}$

$$\Gamma_n^{(k)} \propto [\lambda^{(k)}_{max}]^n$$

and, as far as the $n \to \infty$ limit is concerned, the pre-asymptotic intervals $l^{(k)}$ contribute only an irrelevant prefactor (unless $l^{(k)}$ happens to be normal to the leading eigendirection of $\tilde{T}^{(k)}$). This method of evaluating sums is familiar from statistical mechanics, whence the designation 'transfer operator'. The analogy is purely formal, but often suggestive.

One could now pull out a sequence of leading eigenvalues $\lambda_{max}^{(k)}$ by brute numerical iteration of $\tilde{T}^{(k)}$ and study their $k \to \infty$ limit; this is the essence of the functional equation techniques, such as those employed in [18]. However, as we shall now show, one can do much better: the characteristic equation $0 = \det(1 - zT)$ for the *exact* transfer operator T is available in closed form. We start by observing that $\det(1 - zT)$ can be expressed in terms of traces of T by the identity

$$\det(1 - zT) = \exp[\operatorname{tr}\log(1 - zT)] = \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n}\operatorname{tr}(T^n)\right). \tag{18}$$

Consider evaluating $\operatorname{tr}(T)$ from finite approximations $\tilde{T}^{(k)}$: $\operatorname{tr}(\tilde{T}^{(1)}) = \tilde{T}_{0,0} + \tilde{T}_{1,1}$, $\operatorname{tr}(\tilde{T}^{(2)}) = \tilde{T}_{00,00} + \tilde{T}_{11,11}, \dots$. Clearly the trace of the exact transfer operator is simply the sum of scalings evaluated at the fix points, $\operatorname{tr}(T) = T_{\overline{0},\overline{0}} + T_{\overline{1},\overline{1}}$. More generally, each closed walk through n entries of T contributes a product of the entries along the walk to $\operatorname{tr}(T^n)$. Each step in such a walk shifts the symbolic label by one index; the trace ensures that the walk closes into a periodic string c. We define t_c to be the product of matrix elements along a cycle c; $\operatorname{tr}(T^n)$ is the sum of all such cycles of length n. For example, in figure 2, the cycle $\overline{100}$ contributes $t_{100} = T_{\overline{001},\overline{010}}T_{\overline{010},\overline{100}}T_{\overline{100},\overline{001}}$ to $\operatorname{tr}(T^3)$. In this case the walk is weighted by the product of derivatives along the cycle, $t_{100} = |f'(x_{100})f'(x_{010})f'(x_{001})|^{-1}$. This is clearly cyclically symmetric, so $t_{100} = t_{010} = t_{001}$. For the binary labelled strange sets the first few traces are given by

$$\operatorname{tr}(T) = t_0 + t_1$$
$$\operatorname{tr}(T^2) = t_0^2 + t_1^2 + 2t_{10}$$
$$\operatorname{tr}(T^3) = t_0^3 + t_1^3 + 3t_{100} + 3t_{101}$$
$$\operatorname{tr}(T^4) = t_0^4 + t_1^4 + 2t_{10}^2 + 4t_{1000} + 4t_{1001} + 4t_{1011}. \tag{19}$$

In general, the nth-order trace picks up contributions from all repeats of prime cycles ($m|n$ means that m is a divisor of n);

$$\operatorname{tr}(T^n) = \sum_{n_p|n} n_p t_p^{n/n_p} \tag{20}$$

so the determinant (18) can be re-expressed in terms of prime cycles:

$$\det(1 - zT) = \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{(z^{n_p}t_p)^r}{r}\right) = \exp\left(\sum_p \ln(1 - z^{n_p}t_p)\right) = \prod_p (1 - z^{n_p}t_p).$$

This is the main result of this section; comparing with the (8) of the last section, we see that the dynamical ζ function is related to the transfer operator by

$$1/\zeta = \det(1 - T) = \prod_p (1 - t_p) \tag{21}$$

(we shall usually absorb z into the transfer operator: $zT \to T$, $z^{n_p}t_p \to t_p$). Glancing back, we see that the derivation is very general, and should work for any average over any strange set which satisfies two conditions: (1) the weight associated with a cycle

is multiplicative along the trajectory; (2) the set is organised in such a way that the nearby points in the symbolic dynamics have similar weights.

We summarise the above derivations of ζ functions by recasting them into a more general form. Let the effect of a d-dimensional deterministic map $f(x)$ (or a Poincaré section of a $d+1$-dimensional flow) on a distribution $\phi_\alpha(x)$ be given by a transfer operator $(\mathscr{L} \circ \phi)_\beta(y)$,

$$\mathscr{L}_{\beta\alpha}(y, x) = \omega_{\beta\alpha}(x)\delta(y - f(x)).$$

Here $\omega_{\beta\alpha}(x)$ is any weight factor multiplicative along the trajectory, and the indices refer to possible extra matrix structure (for example, group elements associated with discrete symmetries [12, 11]). The eigenvalues λ_s of $\mathscr{L}$ are given by the zeros of

$$\det(1 - z\mathscr{L}) = \prod_s (1 - z\lambda_s)$$

and the determinant is related by (18) to the traces

$$\operatorname{tr} \mathscr{L}^n = \sum_\alpha \int dx\,dy\,\delta(x - y)\mathscr{L}^n_{\alpha\alpha}(y, x) = \sum_i^{(n)} \frac{\operatorname{tr}\omega^n(x_i)}{|\det(1 - \mathbf{J}^{(n)}(x_i))|}. \tag{22}$$

We assume that no eigenvalue is marginal and factorise the determinant into the product (10) of expanding eigenvalues $\Lambda_{i,1}$, $\Lambda_{i,2}$, ..., $\Lambda_{i,e}$ and contracting eigenvalues $\Lambda_{i,e+1}$, ..., $\Lambda_{i,d-1}$, $\Lambda_{i,d}$:

$$\frac{1}{|\det(1 - \mathbf{J}^{(n)}(x_i))|} = \frac{1}{|\Lambda_i|} \prod_{a=1}^{e} \frac{1}{1 - 1/\Lambda_{i,a}} \prod_{b=e+1}^{d} \frac{1}{1 - \Lambda_{i,b}}.$$

As in (12), Λ_i is the product of expanding eigenvalues. Expanding $1/(1 - 1/\Lambda_{i,a})$, $1/(1 - \Lambda_{i,b})$ as geometric series, and substituting (22) back into (18), one finds that $\det(1 - z\mathscr{L})$ is given by the infinite product

$$\det(1 - z\mathscr{L}) = \prod_{k_1 k_2 \ldots k_e} \prod_{l_{e+1} \ldots l_{d-1} l_d} \frac{1}{\zeta_{k_1 \ldots l_d}(z)}$$

$$\frac{1}{\zeta_{k_1 \ldots k_e l_{e+1} \ldots l_d}(z)} = \prod_p \det\left(1 - \frac{\omega_p}{|\Lambda_p|} \frac{\Lambda_{p,e+1}^{l_{e+1}} \Lambda_{p,e+2}^{l_{e+2}} \ldots \Lambda_{p,d}^{l_d}}{\Lambda_{p,1}^{k_1} \Lambda_{p,2}^{k_2} \ldots \Lambda_{p,e}^{k_e}} z^{n_p}\right). \tag{23}$$

Here 'det' refers to the $\omega_{\beta\alpha}$ indices, and $\omega_p = \prod_{j=0}^{n_p-1} \omega(f^{(j)}(x_p))$. In the escape rates examples considered above, $\omega_{\alpha\beta} = 1$, and the dynamical ζ function (8) is the first, $|\Lambda_p|$ 'volume' weighted term $\zeta_{0\ldots0}(z)$ in the above infinite product. The other terms determine the non-leading, anisotropy-dependent eigenvalues of $\mathscr{L}$.

The transfer operators T, $\mathscr{L}$, ..., appear in the literature under a variety of names, such as the Frobenius–Perron operator [19], etc. The original Ruelle dynamical zeta function [20] is an example of such an average:

$$\zeta_f(z, \varphi) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \operatorname{tr}(T^n) \qquad \operatorname{tr}(T^n) = \sum_i^{(n)} \prod_{j=0}^{n-1} \varphi(f^{(j)}(x_i)). \tag{24}$$

Here the sum goes over all periodic points x_i of period n, and $\varphi(x)$ is a weighting function corresponding to the average one wishes to evaluate. In this case the 'daughters' of x_m are its pre-images $f^{(-1)}(x_m)$, and $T_{dm} = \varphi(x_d)$. We shall, however, also apply the ζ function technique to other types of transfer operators, such as the renormalisation group scaling functions (see paper II).

As we shall show below, a variety of physically interesting averages are determined by the eigenvalues of the transfer operator. By the above relations, those correspond to the zeros of $1/\zeta$, in agreement with the way we have determined the escape rate in the previous section.

Perhaps it is worth emphasising again that the Euler product formula (21) is an expression for the *exact* transfer operator T. We shall extract its eigenvalues directly from (21) with no recourse to any explicit (and coordinatisation-dependent) eigenfunctions. Our cycle expansions will be dominated by short cycles, but that does not mean that we are using finite covers to approximate the set: by resummation that led to (21) we have already been lifted to the topologically exact $k \to \infty$ strange set. The approximation will consist of approximating the strange set under investigation by 'closeby' Cantor sets with a finite number of already asymptotically exact scales. The finite approximations $\tilde{T}^{(k)}$ were introduced only for reasons of pedagogy: our experience is that computations with the asymptotically exact cycles expression (21) are both quicker and of better convergence than computations that go through sequences of finite matrix estimates such as the Markov diagram approximations of [21].

4. Cycle expansions

How are formulae such as (8) used? We start by computing the lengths and eigenvalues of the shortest cycles. This usually requires some numerical work, such as the Newton method searches for periodic solutions; we shall assume that the numerics is under control, and that *all* short cycles up to given length have been found. It is very important not to miss any short cycles, as in this approach the consequences are catastrophic. The result is a list of cycles like figure 3 or table 1 of paper II.

Now we formally expand the Euler product (21)

$$1/\zeta(z) = \prod_p (1 - t_p) = 1 - \sum_{p_1 p_2 \cdots p_k} t_{p_1 + p_2 \cdots + p_k}$$

$$t_{p_1 + p_2 \cdots + p_k} = (-1)^{k+1} t_{p_1} t_{p_2} \cdots t_{p_k} \tag{25}$$

where the sum goes over all distinct non-repeating combinations of prime cycles. For $k > 1$, $t_{p_1 + p_2 \cdots + p_k}$ are 'pseudo' orbits; they are sequences of shorter orbits that shadow the orbit with the symbol sequence $p_1 p_2 \ldots p_k$ along segments $p_1, p_2, \ldots, p_k$ (see figure 4).

For sufficiently small z (we have absorbed z into the weights by $z^{n_p} t_p \to t_p$ substitution) the sum makes sense as a power series in z. For the binary dynamics (example of figure 2, listing of prime cycles in table 1)

$$\begin{aligned}
1/\zeta = {}&(1 - t_0)(1 - t_1)(1 - t_{10})(1 - t_{100})(1 - t_{101})(1 - t_{1000}) \\
&\times (1 - t_{1001})(1 - t_{1011})(1 - t_{10000})(1 - t_{10001}) \\
&\times (1 - t_{10010})(1 - t_{10011})(1 - t_{10101})(1 - t_{10111}) \cdots
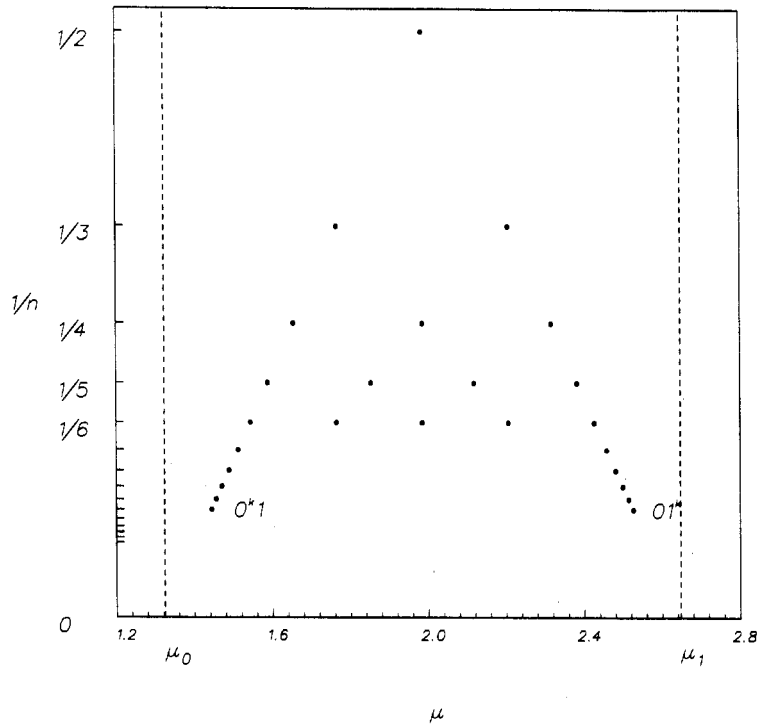\end{aligned} \tag{26}$$

Figure 3. The distribution of the prime cycle eigenvalues for the tent map (30): plotted are the cycle Lyapunov exponents $\lambda_p = \mu_p/n_p \log 2$ against the inverse of the cycle length n_p. The regular structure arises from the factorisation $\Lambda_p = \Lambda_0^{n_0} \Lambda_1^{n_1}$, where n_0 (n_1) is the number of 0s (1s) in the cycle p. In the $1/\zeta$ cycle expansion this infinity of cycles is resummed to the two fundamental cycles, $1/\zeta = 1 - t_0 - t_1$.

the first few terms of the expansion are:

$$
1/\zeta = 1 - t_0 - t_1 - t_{01} - t_{001} - t_{011} - t_{0001} - t_{0011} - t_{0111} - \cdots
$$
$$
- t_{0+1} - t_{0+01} - t_{01+1} - t_{0+001} - t_{0+011} - t_{001+1} - t_{011+1}
$$
$$
- t_{0+01+1} - \cdots \tag{27}
$$

The next step is the key step in our approach: we observe that the expansion (25) allows a regrouping of terms into dominant *fundamental* contributions t_f and decreasing *curvature* corrections c_n:

$$
1/\zeta = 1 - \sum_f t_f - \sum_n c_n. \tag{28}
$$

We shall refer to such series as the *cycle expansions*. For the binary case the cycle expansion is obtained by grouping together the terms of the same total symbol string length (see table 2):

$$
1/\zeta = 1 - t_0 - t_1 - [t_{10} - t_1 t_0] - [(t_{100} - t_{10} t_0) + (t_{101} - t_{10} t_1)]
$$
$$
- [(t_{1000} - t_0 t_{100}) + (t_{1110} - t_1 t_{110}) + (t_{1001} - t_1 t_{001} - t_{101} t_0 + t_{10} t_0 t_1)] - \cdots. \tag{29}
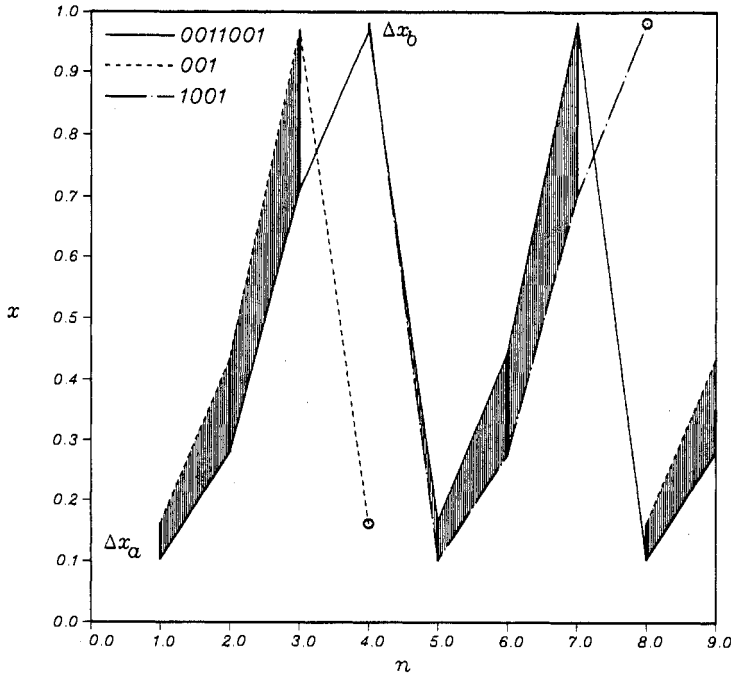$$

Figure 4. Shadowing of $\overline{ab}$ cycle by $\overline{a}$ cycle followed by $\overline{b}$ cycle. Here $a = 001$, $b = 1001$ and $ab = 0011001$. The distance between the cycles is the smallest when the largest number of symbols coincide; then $x_{ab} - x_a \approx 1/\Lambda_a$.

The fundamental cycles t_0, t_1 have no shorter approximants; they are the 'building blocks' of the dynamics in the sense that all longer orbits can be approximately pieced together from them. We call the sum of all terms of the same total length n (grouped in brackets above) the nth curvature correction c_n, for geometrical reasons we shall explain in section 7. If all orbits are weighted equally ($t_p = z^{n_p}$), such combinations cancel exactly; if orbits of similar symbolic dynamics have similar weights, the weights in such combinations will almost cancel.

For example, consider the tent map

$$f_0(x) = \Lambda_0 x \qquad 0 \le x \le 1/\Lambda_0$$
$$f_1(x) = \Lambda_1(1 - x) \qquad 1 - 1/\Lambda_1 \le x \le 1. \tag{30}$$

By the chain rule the stability of any n-cycle factorises as $\Lambda_{\epsilon_1 \epsilon_2 \dots \epsilon_n} = \Lambda_0^m \Lambda_1^{n-m}$, and the stabilities of prime cycles arrange themselves into the regular pattern of figure 3. Clearly the information carried by individual cycles is highly redundant; the cycle expansions, by resumming this infinity of cycles, eliminate the redundancy and extract the generating scales of a strange set. In the example at hand, all curvature terms in (29) vanish (we shall prove this in section 6), and the ζ function is simply

$$1/\zeta = 1 - z/|\Lambda_0| - z/|\Lambda_1|. \tag{31}$$

For strange sets of non–uniform hyperbolicity, the cycle expansions truncated to the fundamental cycles correspond to approximations by strange sets of correct topology but with approximate scales (compare figure 3 with II, figure 5); the curvature corrections account systematically for deviations.

Table 1. Prime cycles for the binary symbolic dynamics up to length 9.

n_p	Cycle	n_p	Cycle	n_p	Cycle	n_p	Cycle	n_p	Cycle
1	0	7	0001001	8	00001111	9	000001101	9	001001111
1	1	7	0000111	8	00010111	9	000010011	9	001010111
2	01	7	0001011	8	00011011	9	000010101	9	001011011
3	001	7	0001101	8	00011101	9	000011001	9	001011101
3	011	7	0010011	8	00100111	9	000100011	9	001100111
4	0001	7	0010101	8	00101011	9	000100101	9	001101011
4	0011	7	0001111	8	00101101	9	000101001	9	001101101
4	0111	7	0010111	8	00110101	9	000001111	9	001110101
5	00001	7	0011011	8	00011111	9	000010111	9	010101011
5	00011	7	0011101	8	00101111	9	000011011	9	000111111
5	00101	7	0101011	8	00110111	9	000011101	9	001011111
5	00111	7	0011111	8	00111011	9	000100111	9	001101111
5	01011	7	0101111	8	00111101	9	000101011	9	001110111
5	01111	7	0110111	8	01010111	9	000101101	9	001111011
6	000001	7	0111111	8	01011011	9	000110011	9	001111101
6	000011	8	00000001	8	00111111	9	000110101	9	010101111
6	000101	8	00000011	8	01011111	9	000111001	9	010110111
6	000111	8	00000101	8	01101111	9	001001011	9	010111011
6	001011	8	00001001	8	01111111	9	001001101	9	001111111
6	001101	8	00000111	9	000000001	9	001010011	9	010111111
6	001111	8	00001011	9	000000011	9	001010101	9	011011111
6	010111	8	00001101	9	000000101	9	000011111	9	011101111
6	011111	8	00010011	9	000001001	9	000101111	9	011111111
7	0000001	8	00010101	9	000010001	9	000110111		
7	0000011	8	00011001	9	000000111	9	000111011		
7	0000101	8	00100101	9	000001011	9	000111101		

Now compare (31) with the Euler product (21). For simplicity take the two scales equal, $|\Lambda_0| = |\Lambda_1| = e^\lambda$. It is a novice error [22] to assume that the infinite Euler product (21) vanishes whenever one of its factors vanishes. If that were true, the factor $(1 - z/|\Lambda_0|)$ would yield

$$0 = 1 - e^{\gamma - \lambda} \tag{32}$$

i.e. the escape rate $\gamma = \ln z$ would equal the stability exponent of the repulsive fixed points. The correct formula follows from (31):

$$0 = 1 - e^{\gamma - \lambda + h} \qquad h = \ln 2. \tag{33}$$

(this is a special case of a general relation between escape rates, Lyapunov exponents and entropies; see section 9). The physical interpretation is that the escape induced by repulsion by each unstable fixed point is diminished by the rate of backscatter from other repelling segments, i.e. the entropy h; the 'false zeros' $z^{n_p} = e^{\lambda_p}$ of the Euler product (21) are shifted by the positive entropy of orbits of the same stability to $z = e^{\lambda - h}$. The fundamental cycles t_f in (28) in this way capture the essential orbit counting and scaling of a dynamical system; as we shall see, the remainder of a cycle expansion are exponentially small curvature corrections.

Given the cycle expansion (29), the calculation is straightforward. We substitute the eigenvalues and lengths of prime cycles (for example, the distinct prime cycles up to length 9 listed in table 1) into the cycle expansion (29), and obtain a polynomial

approximation to $1/\zeta$ (in variable z). The escape rate $\gamma = \ln z$ is determined by finding the leading root of the polynomial approximation. The zeros $1/\zeta(z) = 0$ can be easily determined by standard numerical methods, such as the iterative Newton algorithm

$$q_{n+1} = q_n - \left(\zeta(q_n, \tau) \frac{\partial}{\partial q} \zeta^{-1}(q_n, \tau) \right)^{-1}. \tag{34}$$

Here $z = e^{-q}$ and, if the weights are of form $t_p = e^{\mu_p \tau - \nu_p q}$ (such weights are used in the thermodynamic averages of section 9), the derivative is given by an explicit cycle expansion (no extra numerical work needed), which follows from (25):

$$\frac{\partial}{\partial q} \zeta^{-1}(q, \tau) = - \sum_{p_1 p_2 \dots p_k} \nu_{p_1 + p_2 \dots + p_k} t_{p_1 + p_2 \dots + p_k}$$

$$\nu_{p_1 + p_2 \dots + p_k} = \nu_{p_1} + \nu_{p_2} \dots + \nu_{p_k}. \tag{35}$$

It is easy to check that this cycle expansion also separates into fundamental cycles and curvature corrections.

As we have seen above, the ζ function reduces to a finite polynomial for piecewise linear mappings, but in general the curvature corrections c_n in (28) do not vanish. While the polynomial truncations of the cycle expansions usually already converge well enough, we routinely improve them by fitting $c_2, c_3, \dots, c_N$ with an exponential $c_n = A(zc)^n$, and summing the tail estimate

$$1/\zeta \approx 1 - \sum_f t_f - \sum_{n=n_{\min}}^{N} c_n - \frac{A(zc)^{N+1}}{1 - zc}. \tag{36}$$

We shall justify such exponential tail estimates in section 7; this particular estimate works if the leading pole is real. Alternatively, one can work with cycle expansions of Selberg products, as discussed in II, section 4 and in [14]. Tail resummations often significantly improve the accuracy of the leading root in the cycle expansion; convergence can be further accelerated by Padé approximants [23] or other acceleration techniques [24]. Note also that the existence of a pole at $z = 1/c$ implies that the cycle expansions have a finite radius of convergence, and that analytic continuations will be required for extraction of the non-leading zeros of $1/\zeta$.

A simple illustration of such tail resummation is the ζ function for the Ulam map

$$f(x) = 4x(1 - x) \tag{37}$$

for which the cycle structure is exceptionally simple: the eigenvalue of the $x_0 = 0$ fixed point is 4, while the eigenvalue of any other n-cycle is $\pm 2^n$. Typical cycle weights used in thermodynamic averaging are $t_0 = 4^\tau z$, $t_1 = t = 2^\tau z$, $t_p = t^{n_p}$ for $p \neq 0$. The simplicity of the cycle eigenvalues enables us to evaluate the ζ function by a simple trick: we note that if the value of any n-cycle eigenvalue were t^n, (31) would yield $1/\zeta = 1 - 2t$. There is only one cycle, the x_0 fixed point, that has a different weight $(1 - t_0)$, so we factor it out, multiply the rest by $(1 - t)/(1 - t)$, and obtain a rational ζ function

$$1/\zeta(z) = \frac{(1 - 2t)(1 - t_0)}{(1 - t)}. \tag{38}$$

Consider how we would have detected the pole at $z = 1/t$ without the above trick. As the $\bar{0}$ fixed point is isolated in its stability, we would have kept the factor $(1 - t_0)$ in (26) unexpanded, and noted that all curvature combinations in (29) which include the t_0 factor are unbalanced, so that the cycle expansion is an infinite series:

$$\prod_p (1 - t_p) = (1 - t_0)(1 - t - t^2 - t^3 - t^4 - \ldots) \tag{39}$$

(we shall return to such infinite series in the next section and in II, section 6). The geometric series in the brackets sums up to (38). Had we expanded the $(1 - t_0)$ factor, we would have noted that the ratio of the successive curvatures is exactly $c_{n+1}/c_n = t$; summing as in (38) we would recover the rational ζ function (36).

Table 2. The binary cycle expansion (29) up to length 6, listed in such way that the sum of terms along the pth horizontal line (with the exception of the $t_{100101} + t_{100110}$ pair) is the curvature c_p associated with a prime cycle p.

$-t_0$			
$-t_1$			
$-t_{10}$	$+ t_1 t_0$		
$-t_{100}$	$+ t_{10} t_0$		
$-t_{101}$	$+ t_{10} t_1$		
$-t_{1000}$	$+ t_{100} t_0$		
$-t_{1001}$	$+ t_{100} t_1$	$+ t_{101} t_0$	$- t_1 t_{10} t_0$
$-t_{1011}$	$+ t_{101} t_1$		
$-t_{10000}$	$+ t_{1000} t_0$		
$-t_{10001}$	$+ t_{1001} t_0$	$+ t_{1000} t_1$	$- t_0 t_{100} t_1$
$-t_{10010}$	$+ t_{100} t_{10}$		
$-t_{10101}$	$+ t_{101} t_{10}$		
$\cdot\cdot t_{10011}$	$+ t_{1011} t_0$	$+ t_{1001} t_1$	$- t_0 t_{101} t_1$
$-t_{10111}$	$+ t_{1011} t_1$		
$-t_{100000}$	$+ t_{10000} t_0$		
$-t_{100001}$	$+ t_{10001} t_0$	$+ t_{10000} t_1$	$- t_0 t_{1000} t_1$
$-t_{100010}$	$+ t_{10010} t_0$	$+ t_{1000} t_{10}$	$- t_0 t_{100} t_{10}$
$-t_{100011}$	$+ t_{10011} t_0$	$+ t_{10001} t_1$	$- t_0 t_{1001} t_1$
$-t_{100101}$	$- t_{100110}$	$+ t_{10010} t_1$	$+ t_{10110} t_0$
	$+ t_{10} t_{1001}$	$+ t_{100} t_{101}$	$- t_0 t_{10} t_{101} - t_1 t_{10} t_{100}$
$-t_{101110}$	$+ t_{10110} t_1$	$+ t_{1011} t_{10}$	$- t_1 t_{101} t_{10}$
$-t_{100111}$	$+ t_{10011} t_1$	$+ t_{10111} t_0$	$- t_0 t_{1011} t_1$
$-t_{101111}$	$+ t_{10111} t_1$		

We conclude this section by a comment on the fine structure of curvatures. A glance at the low-order curvatures in table 2 leads to a temptation of associating curvatures with individual cycles, such as $c_{0001} = t_{0001} - t_0 t_{001}$. Numerically such combinations tend to be numerically small (see for example paper II, table 1). Reference [23] goes partially toward fine graining of the curvatures by associating with longer cycles sets of diagrammatically motivated 'complexes'. However, splitting c_n into individual cycle curvatures does not seem possible in general; the first example of such ambiguity in the binary cycle expansion (29) is given by the $\{\overline{001011}, \overline{010011}\}$ $0 \leftrightarrow 1$ symmetric pair of 6-cycles; the counterterm $t_{001} t_{011}$ is shared by the two cycles (see table 2).

5. Pruning

The splitting of cycles into the fundamental cycles and the curvature corrections depends on balancing long cycles t_{ab} against their pseudo-trajectory shadows $t_a t_b$. If the $\overline{ab}$ cycle or either of the shadows $\overline{a}$, $\overline{b}$ do not exist, such curvature cancellation is unbalanced. In a generic dynamical system not every symbol sequence is realised as a physical trajectory; as one looks further and further, one discovers more and more rules which prohibit families of cycles, with unbalanced curvatures of any length, and consequently the cycle expansions are *not* expected to have significantly better convergence than averages computed from sums over covers. Hence the key to a theory of a chaotic dynamical system is firm control of the qualitative, topological enumeration of its possible motions, or the *symbolic dynamics* of the system. In the above we have used examples for which all possible orbits can be labelled by all possible binary sequences. In this section we discuss briefly the cycle expansions for systems with more complicated symbolic dynamics.

A symbolic dynamics is constructed by partitioning the phase space into topologically distinct regions, associating with each region a symbol from an *alphabet*, and using those symbols to label every possible trajectory. *Covering symbolic dynamics* assigns a distinct label to each distinct trajectory: however, there might be symbol sequences which correspond to no trajectory. If all possible symbol sequences can be realised as physical trajectories, the symbolic dynamics is called *complete*; if some sequences are not allowed, the symbolic dynamics is *pruned* (the word is suggested by 'pruning' of branches corresponding to forbidden sequences for symbol dynamics organised by a hierarchical tree). In that case the alphabet must be supplemented by a set of pruning rules, which we shall refer to as the pruning *grammar*.

Symbolic dynamics of a generic dynamic system is arbitrarily complex; even for the logistic map the grammar is finite only for special parameter values. Our strategy is akin to bounding a real number by a sequence of rational approximants; we converge toward the strange set under investigation by a sequence of self-similar Cantor sets. A 'self-similar' Cantor set (in the sense in which we use the word here) is a Cantor set equipped with a *subshift of finite type* [3, 25] symbol dynamics, i.e. the corresponding grammar [26, 27] can be stated as a finite number of pruning rules, each forbidding a finite subsequence $_\epsilon_1 \epsilon_2 \ldots \epsilon_n_-$. Here the notation $_\epsilon_1 \epsilon_2 \ldots \epsilon_n_-$ stands for n consecutive symbols ϵ_1, ϵ_2, ..., ϵ_n, preceded and followed by arbitrary symbol strings. In practice we often find it most expedient to retain the binary labelling and prune directly by setting $t_{\_s\_} = 0$ for any cycle which contains a forbidden substring s. However, the cycle expansions are more efficient if the pruning rules are implemented by redefining the alphabet, as we shall now show by a few examples. They are only illustrative, and the reader is referred to the theory of formal languages textbooks [26–28] for more comprehensive treatment.

Example 1. Alphabet $\{0, 1\}$, prune $\_00\_$.
The pruning rule implies that '0' must always be bracketed by '1's; in terms of a new symbol $2 = 10$, the dynamics becomes unrestricted symbolic dynamics with alphabet $\{1, 2\}$. The cycle expansion (26) becomes

$$
\begin{aligned}
1/\zeta &= (1 - t_1)(1 - t_2)(1 - t_{12})(1 - t_{112}) \ldots \\
&= 1 - t_1 - t_2 - (t_{12} - t_1 t_2) - (t_{112} - t_{12} t_1) - (t_{122} - t_{12} t_2) \ldots \\
&= 1 - t_1 - t_{10} - (t_{110} - t_1 t_{10}) - (t_{1110} - t_{110} t_1) - (t_{11010} - t_{110} t_{10}) \ldots . \quad (40)
\end{aligned}
$$

Table 3. The infinite sequences of cycles used in computing the curvature corrections for binary dynamics with the $\bar{0}$ fixed point pruned. The left-hand side is labelled by the integer labels of equation (42), $\hat{t}_{a_1 \ldots a_k} = \sum_{a=a_k}^{\infty} t_{a_1 \ldots a_{k-1} a}$; the right-hand side by the corresponding binary labels. In the application of II, section 6, the left-hand side labels are the corresponding continued-fraction entries, and the right-hand side are the corresponding binary Farey labels.

n =	1	2	3	4	5	6	7
$\hat{t}_1$ =	t_1 +	t_{10} +	t_{100} +	t_{1000} +	t_{10000} +	t_{100000} +	$t_{1000000}$
$\hat{t}_{12}$ =			t_{110} +	t_{1100} +	t_{11000} +	t_{110000} +	$t_{1100000}$
$\hat{t}_{112}$ =				t_{1110} +	t_{11100} +	t_{111000} +	$t_{1110000}$
$\hat{t}_{23}$ =					t_{10100} +	t_{101000} +	$t_{1010000}$
$\hat{t}_{122}$ =					t_{11010} +	t_{110100} +	$t_{1101000}$
$\hat{t}_{1112}$ =					t_{11110} +	t_{111100} +	$t_{1111000}$
$\hat{t}_{132}$ =						t_{110010} +	$t_{1100100}$
$\hat{t}_{1122}$ =						t_{111010} +	$t_{1110100}$
$\hat{t}_{11112}$ =						t_{111110} +	$t_{1111100}$
$\hat{t}_{34}$ =							$t_{1001000}$
$\hat{t}_{142}$ =							$t_{1100010}$
$\hat{t}_{223}$ =							$t_{1010100}$
$\hat{t}_{1213}$ =							$t_{1101100}$
$\hat{t}_{1132}$ =							$t_{1110010}$
$\hat{t}_{1222}$ =							$t_{1101010}$
$\hat{t}_{11122}$ =							$t_{1111010}$
$\hat{t}_{11212}$ =							$t_{1110110}$
$\hat{t}_{111112}$ =							$t_{1111110}$

This symbolic dynamics describes, for example, circle maps with the golden mean winding number [29].

Example 2. Alphabet $\{0, 1\}$, prune n repeats of '0' $\_000 \ldots 00\_$.
This is equivalent to the n symbol alphabet $\{1, 2, \ldots, n\}$ unrestricted symbolic dynamics, with symbols corresponding to the possible $10 \ldots 00$ block lengths: $2 = 10$, $3 = 100$, ..., $n = 100 \ldots 00$. The cycle expansion (26) becomes

$$1/\zeta = 1 - t_1 - t_2 \ldots - t_n - (t_{12} - t_1 t_2) \ldots - (t_{1n} - t_1 t_n) \ldots . \qquad (41)$$

Example 3. Alphabet $\{0, 1\}$, prune only the fixed point $\bar{0}$.
This is equivalent to the *infinite* alphabet $\{1, 2, 3, 4, \ldots\}$ unrestricted symbolic dynamics. The prime cycles are labelled by all non-repeating sequences of integers, ordered lexically: t_n, $n > 0$; t_{mn}, $t_{mmn}, \ldots, n > m > 0$; $t_{mnr}, r > n > m > 0, \ldots$ (see table 3). Now the number of fundamental cycles is infinite as well:

$$1/\zeta = 1 - \sum_{n>0} t_n - \sum_{n>m>0} \left(t_{mn} - t_n t_m \right) - \sum_{n>m>0} \left(t_{mmn} - t_m t_{mn} \right) - \sum_{n>m>0} \left(t_{mnn} - t_{mn} t_n \right)$$

$$- \sum_{r>n>m>0} \left(t_{mnr} + t_{mrn} - t_{mn} t_r - t_{mr} t_n - t_m t_{nr} + t_m t_n t_r \right) \ldots . \qquad (42)$$

We have already encountered this sum in the Ulam map cycle expansion in section 4. As we shall see in II, section 6, this grammar plays an important role in description of fixed points of marginal stability (see also [9, 30]).

Example 4. Alphabet $\{a, b, c\}$, prune $\_ab\_$.

The pruning rule implies that any string of 'b's must be preceded by a 'c'; so one possible alphabet is $\{a, cb^k; \bar{b}\}$, $k = 0, 1, 2 \ldots$. As the rule does not prune the fixed point $\bar{b}$, it is explicitly included in the list. The cycle expansion (26) becomes

$$1/\zeta = (1 - t_a)(1 - t_b)(1 - t_c)(1 - t_{cb})(1 - t_{ac})(1 - t_{cbb}) \cdots$$

$$= 1 - t_a - t_b - t_c + t_a t_b - (t_{cb} - t_c t_b) - (t_{ac} - t_a t_c) - (t_{cbb} - t_{cb} t_b) \ldots. \quad (43)$$

The effect of the $\_ab\_$ pruning is essentially to unbalance the 2-cycle curvature $t_{ab} - t_a t_b$; the remainder of the cycle expansion retains the curvature form.

Example 5. Alphabet $\{0, 1\}$, prune $\_1000\_, \_00100\_, \_01100\_$.
This example is motivated by the pruning front description of the symbolic dynamics for the Hénon-type maps [31], but that is of no importance here; we offer it as an illustration of a typical pruning sequence.

Step 1. $\_1000\_$ prunes all cycles with a $\_000\_$ subsequence with the exception of the fixed point $\bar{0}$; hence we factor out $(1 - t_0)$ explicitly, and prune $\_000\_$ from the rest. Physically this means that x_0 is an isolated fixed point - no cycle stays in its vicinity for more than two iterations. In the notation of example 2, the alphabet is $\{1, 2, 3; \bar{0}\}$, and the remaining pruning rules have to be rewritten in terms of symbols $2 = 10$, $3 = 100$.

Step 2. Alphabet $\{1, 2, 3; \bar{0}\}$, prune $\_33\_, \_213\_, \_313\_$. Physically the 3-cycle $\bar{3} = \overline{100}$ is pruned and no long cycles stay close enough to it for a single $\_100\_$ repeat. As in example 1, prohibition of $\_33\_$ is implemented by dropping the symbol '3' and extending the alphabet by the allowed blocks 13, 23.

Step 3. Alphabet $\{1, 2, \underline{13}, \underline{23}; \bar{0}\}$, prune $\_213\_, \_23\,13\_, \_13\,13\_$, where $\underline{13} = 13$, $\underline{23} = 23$ are now used as single letters. Pruning of the repetitions $\_13\,13\_$ (the 4-cycle $\overline{13} = \overline{1100}$ is pruned) yields the following.

Result. Alphabet $\{1, 2, \underline{23}, \underline{113}; \bar{0}\}$, unrestricted 4-ary dynamics. The other remaining possible blocks $\_213\_, \_2313\_$ are forbidden by the rules of step 3. The cycle expansion is given by

$$1/\zeta = (1 - t_0)\left(1 - t_1 - t_2 - t_{23} - t_{113} - \sum c_n\right) \quad (44)$$

where c_n are curvature combinations, easily obtained by expanding the 4-ary Euler product.

Example 6. Alphabet $\{0, 1\}$, prune $\_1000\_, \_00100\_, \_01100\_, \_10011\_$.
This somewhat random example of pruning was used in [31], and we shall use it again in II, section 5, to compare the cycle expansions with the more traditional methods. The first three pruning rules were incorporated in the previous example; the last pruning rule $\_10011\_$ leads (in a way similar to example 4) to the alphabet $\{\underline{21^k}, \underline{23}, \underline{21^k 113}; \bar{1}, \bar{0}\}$, and the cycle expansion

$$1/\zeta = (1 - t_0)\left(1 - t_1 - t_2 - t_{23} + t_1 t_{23} - t_{2113} - \sum c_n\right). \quad (45)$$

This concludes our list of examples.

The most important lesson of the pruning of the cycle expansions is that prohibition of a finite subsequence unbalances the head of a cycle expansion and increases the

number of the fundamental cycles in (28). Hence the pruned expansions are expected to start converging only *after* all fundamental cycles have been incorporated—in the last example, the cycles $\overline{1}$, $\overline{10}$, $\overline{10100}$, $\overline{1011100}$ (this is illustrated in II, figures 8, 10). Before the introduction of cycle expansions, no such crisp and clear-cut definition of the fundamental set of scales was available.

If the dynamics is invariant under interchanges of symbols, the symmetry leads to factorisations of the ζ functions and significant simplifications and further improvements of the convergence of the cycle expansions. Such factorisations are discussed in [11].

6. Counting cycles

In this section we shall develop the simplest application of the cycle expansions: the cycle counting, or evaluation of the topological entropies. This information is useful in checking the cycle expansions.

The number of periodic points of length n is given by

$$N_n = \sum_i^{(n)} 1$$

and the corresponding generating function (6) is given by

$$\Omega(z) = \sum_{n=1}^{\infty} z^n N_n.$$

Hence the periodic points are counted by simply setting $t_p = z^{n_p}$ if cycle p exists, $t_p = 0$ if p is pruned. The growth of the number of orbits as a function of the symbol string length is characterised by the *topological entropy*:

$$h = \lim_{n \to \infty} \frac{\log N_n}{n}. \tag{46}$$

By the arguments of section 2, h can be determined from the leading zero $z = e^{-h}$ of the *topological* ζ function [3, 32]

$$1/\zeta = \prod_p (1 - z^{n_p}) = 1 - \sum_f t_f. \tag{47}$$

We emphasise that this expression for the entropy is *exact*; in contrast to the definition (46), no $n \to \infty$ extrapolations of $\ln N_n / n$ are required.

Note that for the cycle counting both t_{ab} and the pseudo-orbit combination $t_{a+b} = t_a t_b$ in (25) have the same value $z^{n_a+n_b}$, so all curvature combinations $t_{ab} - t_a t_b$ vanish exactly, and (47) offers a a quick way of checking the fundamental part of a cycle expansion. If the number of t_f is finite, we refer to this cycle expansion as the *topological polynomial*.

6.1. Counting prime cycles

Our first objective is to evaluate the maximum number of prime cycles M_n for a dynamical system whose symbolic dynamics is built from N symbols. The problem of finding M_n is classical in combinatorics [34, 35] (counting necklaces made out of n beads out of N different kinds) and is easily solved. There are N^n possible distinct strings length n composed of N letters. These N^n strings include all M_d prime d-cycles whose period d equals or divides n. A prime cycle is a non-repeating symbol string: for example, $p = \overline{011} = \overline{101} = \overline{110} = \dots 011011\dots$ is prime, but $\overline{0101} = 010101\dots = \overline{01}$ is not. A prime d-cycle contributes d strings to the sum of all possible strings, one for each cyclic permutation. The total number of possible symbol sequences of length n is therefore related to the number of prime cycles by

$$N^n = \sum_{d|n} d M_d. \tag{48}$$

The number of *prime* cycles follows by Möbius inversion

$$M_n = n^{-1} \sum_{d|n} \mu\left(\frac{n}{d}\right) N^d. \tag{49}$$

where the Möbius function $\mu(1) = 1$, $\mu(n) = 0$ if n has a squared factor, and $\mu(p_1 p_2 \dots p_k) = (-1)^k$ if all prime factors are different.

For example, from two symbols 0, 1 one can form $M_n = 2, 1, 2, 3, 6, 9, 18, 30, 56, 99\dots$ prime cycles, i.e. there are two fixed points $\overline{0}$ and $\overline{1}$, one 2-cycle $\overline{10}$, two 3-cycles $\overline{100}$ and $\overline{101}$, three prime 4-cycles $\overline{1000}, \overline{1001}, \overline{1011}$, etc (see table 1). Similarly, there are $M_n = 3, 3, 8, 18, 48, 116, 312, 810, \dots$ prime cycles built from three symbols, $M_n = 4, 6, 20, 60, 204, 669, 2340\dots$ prime cycles built from four symbols, and so forth.

6.2. Evaluation of topological entropy

Counting cycles amounts to giving each (allowed) prime cycle p weight $t_p = z^{n_p}$ and expanding the Euler product (47) as a power series in z. As all M_n prime cycles of length n have the same weight, the ζ function reduces to

$$1/\zeta(z) = \prod_{n=1}^{\infty} (1 - z^n)^{M_n} = 1 - \sum_{k=1}^{\infty} c_k z^k. \tag{50}$$

For complete symbolic dynamics of N symbols, the cycle expansion is given simply by

$$1/\zeta(z) = \prod_{n=1}^{\infty} (1 - z^n)^{M_n} = 1 - Nz. \tag{51}$$

This follows from the relation of the partition function, in this case

$$\Omega(z) = \sum_{n=1}^{\infty} z^n N^n = \frac{Nz}{1 - Nz}$$

to the ζ function, $\Omega(z) = -z(\partial/\partial z)\log(\zeta)$. Hence for the complete N-ary symbolic dynamics, the topological entropy equals $h = \log N$, as is already clear from the definition (46).

Finiteness of topological polynomials has important implications. In the expansion (28) each curvature c_n is the sum of the M_n prime n-cycle contributions plus or minus 'counterterms' given by products of lower-order cycles: for example, for the binary symbolic dynamics a typical term is

$$c_4 = t_{0001} + t_{0011} + t_{0111} + t_0 t_{01} t_1 - t_0 t_{001} - t_0 t_{011} - t_{001} t_1 - t_{011} t_1. \qquad (52)$$

Note that the half of the contributions to c_k are of negative sign; indeed, if the topological entropy is determined by a finite polynomial of order k, then in the cycle-counting expansion ($t_p = z^{n_p}$) all c_n for $n > k$ must vanish. That means that the first k terms in the cycle expansion are necessary to correctly count the pieces of the Cantor set generated by the dynamical system; they are topologically the *fundamental* cycles of the strange set. It is only after these terms have been included that the cycle expansion is expected to converge smoothly, i.e. only for $n > k$ are the curvatures c_n a measure of the variation of the quality of a linearised covering of the dynamical Cantor set by n-cycle eigenvalues, and expected to fall off rapidly with n. Conversely, if the dynamics is not of a finite subshift type, there is no finite topological polynomial, there are no 'curvature' corrections, and the convergence of the cycle expansions will be poor.

The entropy polynomial also provides useful checks on the correctness of the curvature expansion. For example, observe that in (52) 2^3 terms contribute to c_4, and exactly half of them appear with a negative sign. Such counting rules arise from the identity

$$\prod_p (1 + t_p) = \prod_p \frac{1 - t_p^2}{1 - t_p}. \qquad (53)$$

Substituting $t_p = z^{n_p}$ and using (51) we obtain

$$\prod_p^\infty (1 + z^{n_p}) = \frac{1 - Nz^2}{1 - Nz} = 1 + Nz + \sum_{k=2}^\infty z^k (N^k - N^{k-1}).$$

The z^n coefficient in the above expansion is the number of terms contributing to c_n curvature, so we find that for a complete symbolic dynamics of N symbols and $n > 1$, the number of terms contributing to c_n is $(N - 1)N^{k-1}$ (of which half carry a minus sign).

This technique can be generalised to counting subsets of cycles. Consider the simplest example of a dynamical system with a complete binary tree, a repeller map (30) with two straight branches, which we label 0 and 1. Every cycle weight for such map factorises, with a factor t_0 for each 0, and factor t_1 for each 1 in its symbol string. The transfer matrix traces (19) collapse to $\text{tr}(T^k) = (t_0 + t_1)^k$, and $1/\zeta$ is simply

$$\prod_p (1 - t_p) = 1 - t_0 - t_1. \qquad (54)$$

Substituting into (53) we obtain

$$\prod_p (1 + t_p) = \frac{1 - t_0^2 - t_1^2}{1 - t_0 - t_1} = 1 + t_0 + t_1 + \frac{2 t_0 t_1}{1 - t_0 - t_1}$$

$$= 1 + t_0 + t_1 + \sum_{n=2}^\infty \sum_{k=1}^{n-1} 2 \binom{n-2}{k-1} t_0^k t_1^{n-k}. \qquad (55)$$

Hence for $n \geq 2$ the number of terms in the expansion (50) with k 0s and $n - k$ 1s in their symbol sequences is $2 \binom{n-2}{k-1}$. This is the degeneracy of distinct cycle eigenvalues in figure 3; for systems with non-uniform hyperbolicity this degeneracy is lifted (see II, figure 5).

In order to count the number of prime cycles in each such subset we denote with $M_{n,k}$ ($n = 1, 2, \ldots$; $k = \{0, 1\}$ for $n = 1$; $k = 1, \ldots, n - 1$ for $n \geq 2$) the number of prime n-cycles whose labels contain k zeros, use binomial string counting and Möbius inversion and obtain

$$M_{1,0} = M_{1,1} = 1$$

$$nM_{n,k} = \sum_{m \mid \frac{n}{k}} \mu(m) \binom{n/m}{k/m} \qquad n \geq 2, \ k = 1, \ldots, n - 1$$

where the sum is over all m which divide both n and k.

6.3. Counting pruned cycles

In general, not all prime periodic symbol strings are realised as physical orbits: the M_n calculated above are only an upper bound to the actual number of prime n-cycles. The correct counting requires that the forbidden orbits are pruned, as discussed in section 5. Pruning of the forbidden cycles amounts to setting $t_p = 0$ in (50) for each forbidden sequence p.

The simplest example of pruning is the 'golden mean' pruning, defined in example 1 of section 5. The fundamental cycles in (40) are of length 1 and 2, so the topological polynomial is simply

$$\prod_p (1 - z^{n_p}) = 1 - z - z^2 \tag{56}$$

and the entropy is $h = \log(1 + \sqrt{5})/2$.

Example 4 of section 5 with the alphabet $\{a, cb^k; \bar{b}\}$ is more interesting. In the cycle counting case, the dynamics in terms of $a \to z$, $cb^k \to z/(1 - z)$ is a complete binary dynamics (with the explicit fixed point factor $(1 - t_b) = (1 - z)$):

$$1/\zeta = (1 - z)\left(1 - z - \frac{z}{1 - z}\right) = 1 - 3z + z^2.$$

The topological polynomial for example 6 of section 5

$$1/\zeta = (1 - z)(1 - z - z^2 - z^5 + z^6 - z^7) \tag{57}$$

yields the exact value of the entropy $h = 0.522\ 737\ 642\ldots$, in agreement with the numerical results of [31]. Further examples of topological polynomials for pruned symbolic dynamics are discussed in [11]; there it is also shown that symmetries can lead to factorisations of topological polynomials (and the ζ functions in general).

We conclude this section with a general comment: going from N^n periodic points of length n to M_n prime cycles reduces the number of computations from N^n to $M_n \approx N^{(n-1)}/n$. Use of symmetries can reduce the number of computations by another constant factor [11]. While the resummation of the theory from the partition sum (6) to the cycle expansion (50) thus does not eliminate the exponential growth in the number of orbits, in practice only the short orbits are used, and for them the labour saving is dramatic.

7. Curvatures and nonlinearity

In this section we interpret the curvatures c_n as variations of the weighting function across a flow, and estimate the radius of convergence of cycle expansions by locating the leading pole of $1/\zeta$.

As we have seen above, the 'curvature' terms in the cycle expansion (28) are built up from combinations of form $t_{ab} - t_a t_b$, where $a = \epsilon_1 \epsilon_2 \ldots \epsilon_k$, $b = \epsilon_{k+1} \epsilon_{k+2} \ldots \epsilon_n$ are shorter cycles that shadow the cycle $ab = \epsilon_1 \epsilon_2 \ldots \epsilon_n$ along parts of its orbit (see figure 4). If the cycle weights are multiplicative, as in (24)

$$t_{ab} = \varphi_a(x_{ab})\varphi_b(x_{ba}) \qquad \varphi_a(x_a) = \prod_{j=0}^{k-1} \varphi(x_j) \qquad x_j = f^{(j)}(x_{\epsilon_1 \epsilon_2 \ldots \epsilon_k})$$

the $t_{ab} - t_a t_b$ combination can be estimated from the variation of the weight function $\varphi(x)$ across the gap between $\overline{ab}$ and its shadows $\overline{a}$, $\overline{b}$. Writing the curvature as

$$t_{ab} - t_a t_b = t_{ab} \left[1 - \exp\left(\log \frac{t_a t_b}{t_{ab}} \right) \right] \tag{58}$$

and using

$$\log \frac{\varphi(y)}{\varphi(z)} = \int_{\varphi(z)}^{\varphi(y)} \frac{\mathrm{d}\varphi}{\varphi} = \int_z^y \frac{\varphi'(x)}{\varphi(x)} \,\mathrm{d}x$$

we can write the logarithm in (58) as an integral

$$
\begin{aligned}
\log \frac{t_a t_b}{t_{ab}} &= \log \frac{\varphi_a(x_a)}{\varphi_a(x_{ab})} + \log \frac{\varphi_b(x_b)}{\varphi_b(x_{ba})} \\
&= \int_{x_{ab}}^{x_a} N_a(x)\,\mathrm{d}x + \int_{x_{ba}}^{x_b} N_b(x)\,\mathrm{d}x
\end{aligned}
\tag{59}
$$

over the *nonlinearity* [36]

$$N(x)\,\mathrm{d}x = \frac{\varphi'(x)}{\varphi(x)}\,\mathrm{d}x. \tag{60}$$

In the above, x_a is the fixed point of $f_a(x)$ (i.e. the periodic point $x_{\epsilon_1 \epsilon_2 \ldots \epsilon_k}$ of the original map), x_b the fixed point of $f_b(x)$, x_{ab} is the $\overline{ab}$-cycle point nearest to x_a, and x_{ba} is the $\overline{ab}$-cycle point nearest to x_b (see figure 5). Intuitively, cycle $\overline{ab}$ owns a region V_{ab} of the strange set of size $1/\Lambda_{ab}$, and the integration intervals in (59) $\Delta x_{ab} = x_a - x_{ab}$, $\Delta x_{ba} = x_b - x_{ba}$ are of order $1/\Lambda_a$, $1/\Lambda_a$ respectively. The mean nonlinearity can be defined by

$$\frac{N_{ab}}{\Lambda_{ab}} \equiv \log \frac{t_a t_b}{t_{ab}} \approx \frac{\varphi'(x_a)}{\varphi(x_a)}\Delta x_{ab} + \frac{\varphi'(x_b)}{\varphi(x_b)}\Delta x_{ba}. \tag{61}$$

The curvature corrections for smooth flows are small because of the exponential shrinking of intervals Δx_{ab}, Δx_{ba} across which the nonlinearity is evaluated. Substituting (61) into (58) we have

$$t_{ab} - t_a t_b = t_{ab} \left[1 - \exp\left(\frac{N_{ab}}{\Lambda_{ab}} \right) \right] \approx -N_{ab} \frac{t_{ab}}{\Lambda_{ab}}. \tag{62}$$
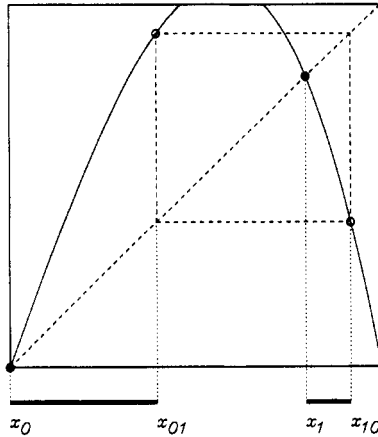
Figure 5. The maps f_a, f_b and the intervals Δx_{ab}, Δx_{ba} used in the evaluation of the nonlinearity (61). The nonlinearity (59) is nearly constant across intervals Δx_{ab}, Δx_{ba} $\approx 1/\Lambda_{ab}$.

As in the limit of no variation of φ this expression tends to zero, N_{ab} is a 'small' constant. N_{ab} is bounded in the following sense; for every pair of cycles $a \neq b$ the cycle expansion (28) contains a series of the form

$$\sum_{k=1}^{\infty} c_{ab^{k-1}} = (t_{ab} - t_a t_b) + (t_{abb} - t_{ab} t_b) + (t_{abbb} - t_{abb} t_b) \ldots + (t_{ab^k} - t_{ab^{k-1}} t_b) \ldots$$

$$\approx -\sum_{k=1}^{\infty} N_{ab^k} \frac{t_{ab^k}}{\Lambda_{ab^k}}.$$

This sequence accumulates toward the b cycle, so $t_{ab^k}/\Lambda_{ab^k} \to (\text{constant}) \times (t_b/\Lambda_b)^k$, and the nonlinearities N_{ab^k} are bounded as they are essentially evaluated at x_b, $N_{ab^k} \to N(x_b)$, so

$$\sum_{k=1}^{\infty} c_{ab^{k-1}} \approx -N(x_b) \frac{1}{1 - t_b/\Lambda_b}. \tag{63}$$

An explicit calculation of such sequence is presented in II, section 3. The cycle expansion contains terms with any block b repeated any number of times, so we expect that the singularities of $1/\zeta$ arise from

$$1/\zeta = 1 - \sum c_n \approx \sum_b N_b \frac{1}{1 - t_b/\Lambda_b}. \tag{64}$$

This sum is of the same structure as the $\Omega(z)$ considered in the derivation of escape rates, with t_b of the partition sum (6) replaced by t_b/Λ_b. This time the hyperbolicity assumption implies that we can neglect the constant prefactors N_b, just as we dropped prefactors a_i from (6). The same chain of arguments that lead from the partition sum (6) to the ζ function now tells us that the poles of (64) should be given by the zeros of

$$1/\zeta_1(z) = \prod_p (1 - z^{n_p} t_p/\Lambda_p). \tag{65}$$

In other words, $1/\zeta(z)$ is expected to have a pole (36)

$$1/\zeta \approx 1 - \sum_f t_f - \sum_{n=n_{\min}}^{N} c_n - \frac{A(z/z_1)^{N+1}}{1 - z/z_1}$$

exactly at the leading zero z_1 of the ζ_1. This can be seen by considering a pair of transfer operators [37]: $\mathscr{L}_0(y,x) = \delta(y - f(x))$, $\mathscr{L}_1(y,x) = (f'(x))^{-1}\delta(y - f(x))$. The ratio of the associated Selberg products (23) yields $1/\zeta$:

$$\frac{\det(1 - z\mathscr{L}_0)}{\det(1 - z\mathscr{L}_1)} = \prod_{k=0}^{\infty} \prod_p \frac{1 - t_p/\Lambda_p^k}{1 - t_p/\Lambda_p^{k+1}} = \frac{1}{\zeta(z)} \tag{66}$$

For nonlinear $f(x)$ the leading eigenvalue $\lambda_0^{(1)}$ of $\det(1 - z\mathscr{L}_1)$ induces a pole in $1/\zeta_0(z)$ and limits the radius of convergence of its cycle expansion (28) to the disk $|z| < |1/\lambda_0^{(1)}|$. The product $\zeta^{-1}\zeta_1^{-1}$ is finite at z_1, with the radius of convergence extended beyond z_1 so that the further zeros of $1/\zeta$ (the next-to-leading eigenvalues of the transfer operator T, section 3) can now be determined. We give an example of such cancellation in the next section, and verify it numerically (and find that it leads to a considerable convergence improvement) in II, section 4. The $1/\zeta_1$ function has a cycle expansion of its own, with curvatures summing up to a pole coinciding with the zero of $1/\zeta_2$, and so forth: this generates an infinite chain of $1/\zeta_k$, whose product is a Selberg [38]-type zeta function

$$Z(z) = \prod_{k=0}^{\infty} 1/\zeta_k(z) = \prod_{k=0}^{\infty} \prod_p (1 - z^{n_p} t_p/\Lambda_p^k) \tag{67}$$

which should have no poles. Indeed, this product is a special case of of $\det(1 - z\mathscr{L})$, equation (23), which, for hyperbolic averages, is an entire function. Note that the sign of the eigenvalue Λ_p is used in this new zeta function, in contrast to the escape rate formulae which depended only on $|\Lambda_p|$. Cycle expansions for such products are discussed in [30] and applied to evaluation of correlation exponents in [14].

The above crude arguments are meant to give the reader a hint of the rich analytic structure of ζ functions, a topic beyond the scope of the present paper; our purpose here is only to sketch how such results can be recovered from cycle expansions. We refer the reader to [7] for rigorous results, and to [23] for the numerical evidence supporting the above guesses about the poles of $1/\zeta$ functions. As far as practical applications of cycle expansions are concerned, such analytic information about the poles and the zeros of ζ functions is very useful because, as we shall show in paper II, it can lead to dramatic improvements in the convergence of cycle expansions.

In conclusion, a curvature contribution $t_{ab} - t_a t_b$ associated with a cycle ab and its shadow cycles a, b is a measure of the inhomogeneity of the weighting functions $\varphi(x)$ across the associated phase-space region $\approx 1/\Lambda_{ab}$. The cycle expansions are expected to converge exponentially, *provided* that the symbolic dynamics is of finite subshift type (so that the contributions of long cycles are always counterbalanced by shorter shadow cycles), that all fundamental cycles are taken into account, and that the non-hyperbolic cycles are excluded from the averages.

8. Stability of a strange set

In this section we introduce the concept of the *stability of a strange set*. It is a part of Sullivan's formulation of the renormalisation theory [39] applied in II, section 4, but we discuss it here because (*a*) it might be a useful general characterisation of strange sets in its own right; (*b*) it offers a simple example of improving the convergence of cycle expansions by exploiting the analytic information about the ζ function.

A strange set (such as the repeller of section 2) is an invariant set of the dynamical flow. The stability of such set can be probed by perturbing infinitesimally its points $x \to x + h(x)$, $h(x) \ll 0$, and investigating the growth of the perturbation under iterations of the mapping. In the nth-level approximation, the strange set can be covered by intervals ℓ_i, as in section 2; we assume that the perturbation $h(x)$ is smooth and essentially constant across small intervals, and replace it by $h_n(x_i)$, where x_i is the periodic point with symbol sequence $\epsilon_1 \epsilon_2 \ldots \epsilon_n$. In one iteration the perturbation $h(x)$ will expand to $f'(x)h(x)$; for any set of 'daughters' $d = \{0\epsilon_1 \ldots \epsilon_n, 1\epsilon_1 \ldots \epsilon_n, \ldots\}$ we define the 'mother' interval perturbation by

$$h_{n-1}(x_m) = f'_0(x_{0m})h_n(x_{0m}) + f'_1(x_{1m})h_n(x_{1m}) + \ldots. \tag{68}$$

By the same argument as in the derivation of escape rates (section 2) we expect the nth iterate perturbation for the entire strange set

$$\Gamma_n = \sum_i^{(n)} \frac{\mathrm{d}f^{(n)}(x_i)}{\mathrm{d}x} \, h_n(x_i) \tag{69}$$

to grow exponentially as $\Gamma_n \propto \delta^n$. We shall introduce here a cycle expansion for estimating the stability eigenvalue δ. Note that the stability defined by the sum (69) is not the stability in the Lyapunov sense (where one would average $\log |f^{(n)\prime}(x_i)|$ over the natural measure, see (90)). The sum (69) tracks the motion of the perturbation's *centre of mass* and keeping the eigenvalue signs in (69) is crucial. The ζ function for this average follows by replacing the transfer matrix (13) by $T_{dm} = f'(x_d)$; it is of the usual Euler product form, with the p–cycle weight given by

$$t_p = z^{n_p} \Lambda_p. \tag{70}$$

For the complete binary symbol dynamics the cycle expansion (29) is given by

$$\frac{1}{\zeta} = 1 - (\Lambda_0 + \Lambda_1)z - (\Lambda_{01} - \Lambda_0\Lambda_1)z^2 - \ldots. \tag{71}$$

The fundamental cycles (here the two fixed points) estimate of the stability is $\delta = 1/z \approx \Lambda_0 + \Lambda_1$, and the curvature corrections are expected to fall off exponentially, as usual. Note however that by (66) the first pole of the cycle expansion (71) is given by the leading zero of $1/\zeta_1 = \prod(1 - t_p/\Lambda_p) = \prod(1 - z^{n_p})$. This, by (51), is simply $1/\zeta_1 = 1 - 2z$, so the first pole of (71) is expected exactly at $z = 1/2$, and should be absent from the cycle expansion of $1/\zeta\zeta_1$:

$$\frac{1 - 2z}{\zeta} = 1 - (\Lambda_0 + \Lambda_1 + 2)z - (\Lambda_{01} - \Lambda_0\Lambda_1 - 2\Lambda_0 - 2\Lambda_1)z^2 + \ldots. \tag{72}$$

By removing the nearby pole at $z = 1/2$, we have 'smoothed out' the cycle expansion and the new fixed point estimate $\delta = 1/z \approx \Lambda_0 + \Lambda_1 + 2$ should be an improvement over the old one; furthermore, the radius of convergence of the cycle expansion has been extended and we expect to be able to determine the next to the leading zero of $1/\zeta$. The numerical work with the fractional map (II, section 5) and the period doubling repeller (II, section 4) supports the above claims.

9. Thermodynamic formalism

In this section we briefly describe the cycle expansion evaluation of some of the thermodynamic averages current in the physics literature. The thermodynamic formalism [4–6] is well known, and we do not intend to review it here: our purpose is to show that once the short cycles and their eigenvalues are known, such averages can be swiftly and elegantly evaluated.

As mentioned in the introduction, the description of a chaotic dynamical system in terms of cycles can be visualised as a tessellation of the dynamical system, with a smooth flow approximated by its periodic orbit skeleton, each region V_i centred on a periodic point x_i, and the size of the region determined by the linearisation of the flow around the periodic point (see figure 1). p_i is the measure of the region V_i associated with the ith periodic point, $p_i = \int_{V_i} d\mu(x)$, so the measure normalisation $\int d\mu(x) = 1$ implies

$$\sum_i^{(n)} p_i = 1. \tag{73}$$

The average of a function $\Phi(x)$ over the strange set is given by

$$\langle \Phi \rangle = \int d\mu(x)\Phi(x) = \lim_{n \to \infty} \sum_i^{(n)} p_i \Phi(x_i). \tag{74}$$

Such averages will be recast here into the cycle expansion form.

In the computation of the asymptotic escape rate (12) from a d-dimensional repeller, each region V_i was given a weight $t_i = z^n |\Lambda_i^{-1}|$, and the escape rate $\gamma = \log z$ determined by balancing the average $1 = \sum t_i$ in the $n \to \infty$ limit. In a class of applications of the thermodynamic formalism one generalises such weights to 'moments' $t_i = z^n |\Lambda_i|^\tau p_i^q$. By varying $z = e^\sigma$ one can investigate the distribution of the cycle lengths; varying τ probes the distribution of cycle stabilities, and varying q explores the measure distribution. In the escape rate example (12), p_i is the 'natural' measure [40, 41]

$$p_i = |\Lambda_i^{-1}| e^{n\gamma} \tag{75}$$

but for the time being we desist from explicitly fixing p_i. The transfer matrix (13) is now replaced by

$$T_{dm} = z |\Lambda_d/\Lambda_m|^\tau (p_d/p_m)^q. \tag{76}$$

With proviso that the measure p_i is also multiplicative along the flow, and that the closeby trajectories have closeby measures, so that the condition (15) is fulfilled, the derivation of section 3 can be repeated *in toto*. The result is a ζ function of the same form (21) as usual, but with the weight t_p replaced by

$$t_i = z^n |\Lambda_i|^\tau p_i^q \to t_p = e^{n_p\sigma + \mu_p\tau - \nu_p q}. \tag{77}$$

Here n_p is the topological length of p; $\sigma = \ln z$; $\mu_p = \ln |\Lambda_p|$ is the stability exponent of p, where Λ_p is the product of expanding eigenvalues; and $\nu_p = -\ln \prod p_k$, where $\prod p_k$ is evaluated along the cycle p. The relations among thermodynamic functions, such as

$q = q(\tau)$ (for constant σ), are determined by solving $0 = \prod(1 - t_p)$. For example, for the 2-scale Cantor set (30) with equipartition measure $p_i = 2^{-n}$, the cycle expansion is

$$0 = 1 - e^{\mu_0 \tau} 2^{-q} - e^{\mu_1 \tau} 2^{-q}$$

so $q(\tau) = \ln(e^{\mu_0 \tau} + e^{\mu_1 \tau})/\ln 2$. Once $q(\tau)$ is determined, its derivatives are available as well, as explicit cycle expansions. For example, a variation in q, τ which respects the $0 = \prod(1 - t_p)$ condition

$$0 = \left(dq \frac{\partial}{\partial q} + d\tau \frac{\partial}{\partial \tau} \right) \prod_p (1 - t_p) \tag{78}$$

together with (35), yields

$$\frac{d\tau}{dq} = \frac{\sum \nu_{p_1 + p_2 \ldots + p_k} t_{p_1 + p_2 \ldots + p_k}}{\sum \mu_{p_1 + p_2 \ldots + p_k} t_{p_1 + p_2 \ldots + p_k}}. \tag{79}$$

This is evaluated by substituting the available short cycles into the denominator and numerator cycle expansions (optionally with the convergence improved by the tail estimates (36)). Similarly,

$$\frac{d^2\tau}{dq^2} = -\frac{\sum (\nu_{p_1 + p_2 \ldots + p_k} - \mu_{p_1 + p_2 \ldots + p_k} (d\tau/dq))^2 t_{p_1 + p_2 \ldots + p_k}}{\sum \mu_{p_1 + p_2 \ldots + p_k} t_{p_1 + p_2 \ldots + p_k}} \tag{80}$$

and so on.

Some familiar examples of such averages are as follows.

Example 1. The generalised dimensions [42]

$$D_q = \frac{1}{q-1} \tau(q) \qquad t_i = \frac{p_i^q}{l_i^\tau} \rightarrow t_p = e^{\mu_p \tau - \nu_p q}. \tag{81}$$

$\tau(q)$ is determined from the cycle expansion as explained in section 4. A plot of a Legendre transform, either [42–44]

$$f(\alpha) = -\tau(q) + q\alpha(q) \qquad \alpha(q) = \frac{d\tau}{dq} \tag{82}$$

or [21, 24]

$$s(\mu) = q(\tau) - \tau \mu(\tau) \qquad \mu(\tau) = \frac{dq}{d\tau} = \frac{1}{\alpha}. \tag{83}$$

is usually more informative than a plot of $\tau(q)$ or D_q. Such functions are plotted by evaluating $q(\tau)$, $q'(\tau)$ from the cycle expansions (29), (79) for a range of τ, and the $\tau \rightarrow \pm\infty$ ends are fixed by investigating the maximal and minimal scales of the strange set (examples are given in paper II).

By the normalisation (73), $q(0) = 1$. For 1D maps, $D_0 = -\tau(0)$ determined by solving

$$0 = \prod_p (1 - |\Lambda_p|^{-D_0}) \tag{84}$$

is the Hausdorff–Besicovitch dimension D_H (irrespective of the choice of measure; provided that the covering intervals l_i are optimal). For example, for piecewise linear sets with subshift of finite type symbol dynamics of section 5 with alphabet $\{a_1, a_2, \ldots, a_f\}$, D_H is determined by [45]

$$0 = 1 - \sum_{i=1}^{f} |\Lambda_{a_i}|^{-D_H}.$$

The information dimension D_1 is given by the cycle expansion (79):

$$D_1 = \frac{d\tau}{dq}\bigg|_{q=1, \tau=0} = \frac{\sum v_{p_1 + p_2 \ldots + p_k} e^{-v_{p_1} + p_2 \ldots + p_k}}{\sum \mu_{p_1 + p_2 \ldots + p_k} e^{-v_{p_1} + p_2 \ldots + p_k}} = \frac{\langle v \rangle}{\langle \mu \rangle} \tag{85}$$

As μ_p is here defined as the sum of the expanding stability exponents, the above definition of the generalised dimensions is the standard one only for the case of a single expanding eigenvalue, in which case D_q is the partial dimension [31, 46–48]. In II, section 5 we apply the above formulae to the evaluation of partial dimensions for the Hénon type attractors, with

$$t_p = e^{\mu_p^{u,s}\tau - v_p q}. \tag{86}$$

where $\mu_p^{u,s} = \ln |\Lambda_p^{u,s}|$ is the stability exponent in the expanding (contracting) direction. References [11, 12] test the corresponding expression for a 2D Hamiltonian pinball model; the convergence of the cycle expansions is as good as expected. However, as shown in [10], even for a repeller as simple as a 2D disconnected fractal generated by a pair of linear maps [49] with only expanding eigenvalues, the cycle expansion for the Hausdorff dimension converges poorly. The reason is that the d-dimensional cover used in the standard definition [7] of the Hausdorff dimension is a ball of radius proportional to the inverse of the least expanding eigenvalue $\Lambda_i^{\min}$; while the Jacobians (11) are multiplicative along the flow, their eigenvalues are not, and the curvature corrections $\Lambda_{ab}^{\min} - \Lambda_a^{\min}\Lambda_b^{\min}$ need not be small.

Example 2. The generalised Kolmogorov entropies [50]

$$K_q = \frac{1}{q-1}\sigma(q) \qquad t_i = p_i^q e^{n\sigma} \rightarrow t_p = e^{n_p \sigma - v_p q}. \tag{87}$$

By the normalisation (73), $q(0) = 1$. $K_0 = -\sigma(0) = h$ is the topological entropy, already discussed in section 6. K_q is constructed in such a way that the K_1 is the metric entropy [51] $K_1 = \lim_{n \to \infty} \sum^{(n)} p_i \log p_i / n$, where $p_i = e^{-v_i}$ is the probability of finding a symbol string i. The K_1 cycle expansion is similar to (79):

$$K_1 = \frac{d\sigma}{dq}\bigg|_{q=1, \sigma=0} = \frac{\sum v_{p_1 + p_2 \ldots + p_k} e^{-v_{p_1} + p_2 \ldots + p_k}}{\sum (n_{p_1} + n_{p_2} \ldots + n_{p_k}) e^{-v_{p_1} + p_2 \ldots + p_k}} = \frac{\langle v \rangle}{\langle n \rangle}. \tag{88}$$

Example 3. The generalised Lyapunov exponent [40, 52]

$$\lambda_\tau = -\frac{\sigma(\tau)}{\tau} \qquad t_i = e^{n\sigma}|\Lambda_i|^\tau p_i \rightarrow t_p = e^{n_p \sigma + \mu_p \tau - v_p}. \tag{89}$$

By the normalisation (73), $\sigma(0) = 0$. The weight is constructed in such a way that the Lyapunov exponent (in d dimensions, the sum of the positive Lyapunov exponents)

$$\langle \lambda \rangle = \int d\mu(x)\, \lambda(x) \approx \frac{1}{n} \sum_{i}^{(n)} p_i \ln |\Lambda_i| \tag{90}$$

(see (74)) is given by

$$\lambda_0 = \frac{d\lambda}{d\tau}\bigg|_{\tau=0,\sigma=0} = \frac{\sum \mu_{p_1+p_2\ldots+p_k}\, e^{-\nu_{p_1+p_2\ldots+p_k}}}{\sum (n_{p_1} + n_{p_2} \ldots + n_{p_k})\, e^{-\nu_{p_1+p_2\ldots+p_k}}} = \frac{\langle \mu \rangle}{\langle n \rangle}. \tag{91}$$

The above three quantities D_1, K_1 and λ_0 satisfy the identity [46]

$$K_1 = \lambda_0 D_1 \tag{92}$$

irrespective of the choice of measure.

The most common choices of measure are the equipartition measure (all symbol sequences of length n weighted equally), and the natural measure (regions of a strange set are weighted according to their visitation frequency).

The *equipartition* (or *cyclinder*) measure presents no problem. The growth of the number of allowed symbol sequences N_n with the sequence length n is characterised by the topological entropy $h = \lim_{n\to\infty} \ln(N_n)/n$ (see section 6) and for the equipartition measure

$$p_i = 1/N_n \to \nu_p = n_p h. \tag{93}$$

The *natural* measure is defined by the long term average

$$p_i = \lim_{N\to\infty} \frac{N_i}{N}. \tag{94}$$

Here N_i is the number of times a 'typical' trajectory of length N visits the ith region of the strange set. This average is problematic and often hard to control in the physically interesting situations, as the generic dynamical systems are neither uniformly hyperbolic nor structurally stable: it is not known whether even the simplest realistic model of a strange attractor, the Hénon attractor [53], is a strange attractor or merely a long stable cycle. One way to circumvent such subtleties is to use the repeller measure (75), computed on the union of unstable orbits, as the working definition of the 'natural' measure.

While the relation (92) holds irrespective of the choice of measure, customarily all of the above generalised thermodynamic exponents are defined with respect to the natural measure. This choice leads to additional relations. For example, for a repeller the natural measure (75) substituted in (88) leads to a relation [40]

$$K_1 = \frac{\langle -n\gamma + \mu \rangle}{\langle n \rangle} = \lambda_0 - \gamma \tag{95}$$

between the metric entropy, the sum of the expanding Lyapunov exponents, and the escape rate. This is the d-dimensional generalisation of the relation (33) discussed in

section 4. On a strange attractor the escape rate $\gamma = 0$, and the above relation reduces to the equality [54] $K_1 = \lambda_0$.

However, $p_i = 1/|\Lambda_i|$ is the natural measure only for the strictly hyperbolic systems [4–6]. For non-hyperbolic systems, the measure develops folding cusps. For example, for Ulam-type maps (unimodal maps with quadratic critical point mapped onto the 'left' unstable fixed point x_0, discussed in more detail in II, section 2), the measure develops a square-root singularity on the $\bar{0}$ cycle:

$$p_0 = \frac{1}{|\Lambda_0|^{1/2}}. \tag{96}$$

The thermodynamics averages are still expected to converge in the 'hyperbolic' phase [31, 41] where the positive entropy of unstable orbits dominates over the marginal orbits, but they fail in the 'non-hyperbolic' phase [47, 55].

One is by no means forced to use either the natural or the equipartition measure; there is a variety of other choices [24], depending on the problem and one's taste. Also the stability Λ_i need not refer to motion in the dynamical space; in more general settings it can be a renormalisation scaling function [56] (II, section 4), or even a scaling function describing a strange set in the parameter space (II, section 6).

To summarise, the cycle expansions generalise smoothly to 'thermodynamic' averages, except for generalisations of dimensions to higher-dimensional flows and delicacies of defining the 'natural' measure for non–hyperbolic flows. These are not problems of cycle expansions, but much deeper problems of the theory of dynamical systems: what is a good characterisation of non-uniform, anisotropic strange sets? In addition, it is not clear whether the problems of the version of the thermodynamic formalism presented above are of physical import. In contrast to somewhat *ad hoc* thermodynamic moments (77) (which have no good generalisation to higher dimensional, non-isotropic flows, and which can be evaluated only by computer manipulation of microscopic measurements of the strange set), our experience is that for the physically motivated averages, such as the escape rates and quantum resonances [11–13], the cycle expansions work well.

10. Summary and conclusions

A motion on a strange attractor can be approximated by shadowing long orbits by sequences of nearby shorter periodic orbits. This notion has here been made precise by approximating orbits by primitive cycles, and evaluating associated curvatures. A curvature measures the deviation of a long cycle from its approximation by shorter cycles; the smoothness of the dynamical system implies exponential fall-off for (almost) all curvatures. We propose that the theoretical and experimental strange sets be presented in terms of the symbol sequences of short cycles (a topological characterisation of the spatial layout of the strange set) and their eigenvalues (metric structure); for example, plotted as figure 3 or listed as in II, table 1. The cycle expansions then offer an efficient method for evaluating periodic orbit averages; accurate estimates can already be obtained from a few fundamental cycles.

For reasons of clarity we have here motivated the cycle expansions by a simple 1D repeller of figure 2; detailed investigations of a series of low-dimensional chaotic systems undertaken in the sequel paper II, as well as the classical and quantum pinball studies of [11–13], give us some confidence in the general feasibility of the cycle analysis

advocated here. The cycle expansions such as (28) outperform the pedestrian methods such as extrapolations from the finite cover sums (2) for a number of reasons. The cycle expansion is a better averaging procedure than the naive box counting algorithms because the strange attractor is here pieced together in a topologically invariant way from neighbourhoods ('space average') rather than explored by a long ergodic trajectory ('time average'). The cycle expansion is co-ordinate and reparametrisation invariant—a finite nth-level sum (2) is not. Cycles are of finite period but infinite duration, so the cycle eigenvalues are already evaluated in the $n \to \infty$ limit, but for the sum (2) the limit has to be estimated by numerical extrapolations. And, crucially, the higher terms in the cycle expansion (28) are deviations of longer primitive cycles from their approximations by shorter cycles. Such combinations vanish exactly in piecewise linear approximations and fall off exponentially for smooth dynamical flows.

However, the cycle expansions are not magic, and they will not converge any better than the more traditional thermodynamic sums *unless the following prerequisites are met*.

(1) The essential prerequisite for implementing the above 'shadowing' is a good understanding of the symbolic dynamics of the dynamical system; the present formulation requires that the symbolic dynamics be of a finite subshift type (see section 5). A generic dynamical system is *not* of that type: our strategy is to approach it by a sequence of finite subshift approximants, just as a generic number can be bracketed by a sequence of rational approximants.

(2) The weight used in averaging must be multiplicative along the flow, and the flow should be smooth, so that nearby trajectories have nearby weights.

(3) Cycle expansions converge only in the hyperbolic phase, i.e. only for averages dominated by the positive entropy of unstable cycles. Marginal fixed points show up indirectly, as power-law corrections and non-analyticities of the ζ functions. If a sequence of fundamental cycles t_f is infinite and accumulating toward marginal stability, the sequence must be summed up in order that the convergence of the cycle expansion be exponential (see II, section 6).

(4) As developed here, the cycle expansions are good only for extracting the leading eigenvalues of transfer operators. If more eigenvalues are needed, techniques for analytically continuing beyond the leading singularities must be developed.

When and if the cycles suffice for the complete characterisation (and reconstruction) of a dynamical system is not known [57], but they do go further toward detailed invariant characterisation of low-dimensional chaotic dynamical systems than other current methods, and we hope that in the future the data will be presented in terms of cycles rather than 'thermodynamic' averages.

References

[1] Poincaré H 1892 *Les méthodes nouvelles de la méchanique céleste* (Paris: Gauthier-Villars)
[2] MacKay R S and Miess J D 1987 *Hamiltonian Dynamical Systems* (Bristol: Hilger)
[3] Smale S 1967 *Bull. Am. Math. Soc.* **73** 747
[4] Sinai Ya G 1972 *Russ. Math. Surveys* **166** 21
[5] Bowen R 1975 *Equilibrium states and the ergodic theory of Anosov-diffeomorphisms* (*Lecture Notes in Mathematics* **470**) (Berlin: Springer)
[6] Ruelle D 1978 *Statistical Mechanics, Thermodynamic Formalism* (Reading, MA: Addison-Wesley)
[7] Ruelle D 1989 *Elements of Differentiable Dynamics and Bifurcation Theory* (New York: Academic)
[8] Cvitanović P 1988 *Phys. Rev. Lett.* **61** 2729
[9] Cvitanović P, Gunaratne G H and Vinson M J 1990 *Nonlinearity* **3** submitted
[10] Aurell E 1990 *Phys. Rev. A* submitted
[11] Cvitanović P and Eckhardt B in preparation
[12] Cvitanović P and Eckhardt B 1989 *Phys. Rev. Lett.* **63** 823
[13] Cvitanović P, Eckhardt B and Scherer P in preparation
[14] Christiansen F, Paladin G and Rugh H H 1990 *Phys. Rev. A* submitted
[15] Kadanoff L and Tang C 1984 *Proc. Natl Acad. Sci. USA* **81** 1276
[16] Ruelle D 1986 *J. Stat. Phys.* **44** 281
     Parry W and Pollicott M 1983 *Ann. Math.* **118** 573
[17] Mandelbrot B B 1983 *The Fractal Geometry of Nature* (San Francisco: Freeman)
[18] Feigenbaum M J, Procaccia I and Tél T 1989 *Phys. Rev. A* **39** 5359
[19] Oono Y and Takahashi Y 1980 *Prog. Theor. Phys.* **63** 1804
     Chang S-J and Wright J 1981 *Phys. Rev. A* **23** 1419
     Takahashi Y and Oono Y 1984 *Prog. Theor. Phys.* **71** 851
[20] See [6], section 7.23
[21] Feigenbaum M J 1987 *J. Stat. Phys.* **46** 919, 925
[22] Voros A 1988 *J. Phys. A: Math. Gen.* **21** 685
[23] Aurell E 1990 *J. Stat. Phys.* to appear
[24] Artuso R, Cvitanović P and Kenny B G 1989 *Phys. Rev. A* **39** 268
[25] Guckenheimer J and Holmes P 1986 *Non-linear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (New York: Springer)
[26] Salomaa A 1973 *Formal Languages* (New York: Academic)
[27] Hopcroft J E and Ullman J D 1979 *Introduction to Automata Theory, Languages, and Computation* (Reading, MA: Addison-Wesley)
[28] Cvetković D M, Doob M and Sachs H 1980 *Spectra of Graphs* (New York: Academic)
[29] Feigenbaum M J 1988 *Nonlinearity* **1** 577
[30] Cvitanović P 1990 *Nonlinear Physical Phenomena, Brasilia 1989 Winter School* ed À Ferraz, F Oliveira and R Osorio (Singapore: World Scientific)
[31] Cvitanović P, Gunaratne G H and Procaccia I 1988 *Phys. Rev. A* **38** 1503
[32] The ζ functions were originally introduced in this context by E Artin and B Mazur (1965 *Ann. Math.* **81** 82). See for example [33] for their evaluation for maps of the interval.
[33] Milnor J and Thurston W 1988 On iterated maps of the interval *Dynamical Systems, Proc., University of Maryland 1986–87* (*Lecture Notes in Mathematics* **1342**) **ed A Dold and B Eckmann** (Berlin: Springer) p465
[34] Riordan J 1958 *An Introduction to Combinatorial Analysis* (New York: Wiley)
     Gilbert E N and Riordan J 1961 *Illinois J. Math.* **5** 657
[35] Brucks K M 1987 *Adv. Appl. Math.* **8** 434
[36] Sullivan D 1987 *Non-linear Evolution and Chaotic Phenomena* ed P Zweifel, G Gallavotti and M Anile (New York: Plenum)
[37] Ruelle D 1976 *Inventiones Math.* **34** 231
[38] Selberg A 1956 *J. Ind. Math. Soc.* **20** 47
[39] Sullivan D 1989 *Universality in Chaos* 2nd edn, ed P Cvitanović (Bristol: Hilger)
[40] Kantz H and Grassberger P 1985 *Physica* D **17** 75
[41] Grebogi C, Ott E and Yorke J A 1987 *Phys. Rev. A* **36** 3522
[42] Grassberger P 1983 *Phys. Lett.* **97A** 227; 1985 *Phys. Lett.* **107A** 101
     Hentschel H G E and Procaccia I 1983 *Physica* D **8** 435
     Benzi R, Paladin G, Parisi G and Vulpiani A 1984 *J. Phys. A: Math. Gen.* **17** 3521
[43] Halsey T C, Jensen M H, Kadanoff L P, Procaccia I and Shraiman B I 1986 *Phys. Rev. A* **107** 1141

[44] Even though the thermodynamic formalism is of older vintage (we refer the reader to [6] for a comprehensive overview), we adhere here to the notational conventions of [43] which are more current in the physics literature.

[45] Moran P A P 1946 *Proc. Camb. Phil. Soc.* **42** 15

[46] Eckmann J-P and Ruelle D 1980 *Rev. Mod. Phys.* **57** 617

[47] Ott E, Grebogi C and Yorke J A 1989 *Phys. Lett.* **135A** 343

[48] Tél T 1989 *J. Phys. A: Math. Gen.* **22** L691

[49] Barnsley M 1988 *Fractals Everywhere* (New York: Academic)

[50] Grassberger P and Procaccia I 1985 *Phys. Rev. A* **31** 1872

[51] Shannon C 1948 *Bell Systems Tech. J.* **27** 379

[52] Fujisaka H 1983 *Prog. Theor. Phys.* **70** 1264

[53] Hènon M 1976 *Commun. Math. Phys.* **50** 69

[54] Pesin Ya B 1977 *Usp. Mat. Nauk* **32** 55 (*Russian Math. Surveys* **32** 55)

[55] Politi A, Badii R and Grassberger P 1988 *J. Phys. A: Math. Gen.* **15** L763
Grassberger P, Badii R and Politi A 1988 *J. Stat. Phys.* **51** 135

[56] Feigenbaum M J 1988 *J. Stat. Phys.* **52** 527

[57] It is argued in [56] that cycles do not suffice for a complete characterisation of strange sets.