Extracting dynamical behaviour via Markov models

Gary Froyland*

Abstract

Statistical properties of chaotic dynamical systems are difficult to estimate reliably. Using long trajectories as data sets sometimes produces misleading results. It has been recognised for some time that statistical properties are often stable under the addition of a small amount of noise. Rather than analysing the dynamical system directly, we slightly perturb it to create a Markov model. The analogous statistical properties of the Markov model often have "closed forms" and are easily computed numerically. The Markov construction is observed to provide extremely robust estimates and has the theoretical advantage of allowing one to prove convergence in the noise $\to 0$ limit and produce rigorous error bounds for statistical quantities. We review the latest results and techniques in this area.

Contents

*Research supported by the Deutsche Forschungsgemeinschaft under Grant De 448/5-4.

1 Introduction and basic constructions

Suppose that we find ourselves presented with a discrete time[1] dynamical system, and we would like to perform some (mainly ergodic-theoretic) analysis of the dynamics. We are not concerned with the problem of embedding, nor with the extraction of a system from time series. We assume that we have been presented with a dynamical system and do not question its validity.

Any analysis of a dynamical system involving *average* quantities requires a reference measure with which to average contributions from different regions of phase space. Often the measure that one wishes to use is the probability measure that is described by the distribution of a typical long trajectory of the system; it is commonly called the *natural measure* or *physical measure* of the system.

1.1 What do we do?

This chapter discusses a method of modelling the dynamics by a finite state Markov chain. Naturally such a model contains much less information than the original dynamical system. However, this simplification of the dynamics allows the exact computation of many properties of the Markov chain which correspond to important indicators and properties of the original dynamics. For example, finding invariant sets, obtaining invariant measures, calculating rates of mixing and the spectrum of eigenvalues of transfer operators, computing means and variances of recurrence times, and estimating Lyapunov exponents; all of these calculations are *exact*[2] for the Markov chain. We hope that although we are throwing away a lot of information in our Markov model, we retain the essential properties of the original system. The questions then are: (i) Are these quantities computed for the Markov model good estimators of the corresponding quantities for the original system, and (ii) how best to define these Markov models for various sorts of systems.

1.2 How do we do this?

We describe the fundamental construction of the modelling process. Consider a dynamical system (M, T) defined by a map $T : M \circlearrowleft$, where M is a compact subset of $\mathbb{R}^d$. Partition the phase space into a finite number of connected sets $\{A_1, \ldots, A_n\}$ with nonempty interior. Usually, this partition will take the form of a regular grid covering the phase space M. We now completely ignore any dynamics that occurs inside each individual partition set, and focus only on the coarse-grained dynamics displayed by the evolution of whole partition sets. To form our Markov model, we identify each set A_i with a state i of our n-state Markov chain. We construct an $n \times n$ transition matrix P, where the entry P_{ij} is to be interpreted as:

$$P_{ij} = \text{ the probability that a typical point in } A_i \text{ moves into } A_j \text{ under one iteration of the map } T. \quad (1)$$

We now meet the notion of *typicality* and begin to impinge on ergodic-theoretic ideas. Leaving formality for the moment, we shall assume that the trajectories $\{x, Tx, T^2x, \ldots\}$ of Lebesgue almost all initial points $x \in M$ have the same distribution on M. This distribution may be represented as a probability measure, denoted by μ. Now, in light of (1), the most

[1] Similar constructions for flows are possible by considering the "time-t" map.

[2] A further approximation must be introduced for the calculation of Lyapunov exponents.

natural definition of P_{ij} is

$$\tilde{P}_{ij} = \frac{\mu(A_i \cap T^{-1}A_j)}{\mu(A_i)} \tag{2}$$

since $\tilde{P}_{ij} = \mathrm{Prob}_\mu\{Tx \in A_j | x \in A_i\}$. Unfortunately, the natural measure μ is usually unknown and for want of a better alternative, we compute a (slightly) different transition matrix using normalised Lebesgue measure m instead of μ:

$$P_{ij} = \frac{m(A_i \cap T^{-1}A_j)}{m(A_i)} \tag{3}$$

Several numerical techniques have been put forward regarding the computation of P; see §6.1.

1.3 Why do we do this?

The alternative to Markov modelling of the dynamics via some coarse graining is to simply simulate very long orbits of the dynamical system. For the purposes of this discussion, we restrict ourselves to the problem of approximating the probability measure μ that describes the asymptotic distribution of almost all trajectories.

Given a long orbit $\{x, Tx, \ldots, T^{N-1}x\}$, one is implicitly approximating the long-term distribution μ by the finite-time distribution $\mu_N(x) := \frac{1}{N}\sum_{k=0}^{N-1}\delta_{T^k x}$, where δ_x is the delta-measure at $x \in M$. This is certainly a simple way to compute an approximate invariant measure (or long-term distribution), as it does not involve any complicated matrix constructions; one just iterates one's map. There are however, drawbacks to this simple approach. It is possible that orbits display transient (non-equilibrium) behaviour for lengthy periods of time before settling into a more regular (statistically speaking) mode. Thus, by following a simulated orbit for a finite time, there is the risk that one is only observing this transient behaviour and not the true asymptotic behaviour of the system. There is also the problem of computer round-off; try to find the long-term distribution of the tent map or the circle doubling map by iterative simulation (all trajectories are attracted to 0 in finite time). These are extreme cases, but the potential compounding inaccuracy of long computer generated orbits should not be forgotten. Let's be generous though, and assume that our approximation $\mu_N(x)$ actually does (weakly) converge to μ. How fast does this happen? What is the convergence rate with respect to the length of the orbit N? Can one produce rigorous error bounds for the difference between the distributions $\mu_N(x)$ and μ? For the most part, the answer to each of these questions is "We don't know yet". In toy cases, one can produce extremely crude probabilistic *lower* bounds for the error, of the form $C/\sqrt{N}$, but this is not really satisfactory.

Our method of Markov modelling attempts to overcome all of these difficulties. Transient effects are completely removed as we model the system as a Markov chain and its asymptotic behaviour may be computed *exactly*. Computer roundoff is not so much of a problem, as we are now only computing a *single iterate* of the dynamics rather than a compounding sequence of iterations. The constructions also permit a rigorous treatment of questions like rates of convergence, error bounds, and even just whether convergence occurs at all. Finally, from the practical point of view, the method of Markov coarse graining is often very efficient computationally, producing better answers in less time.

This discussion has primarily been aimed at the computation of invariant measures, but it applies also to the computation of other dynamical indicators such as the rate of decay of correlations, Lyapunov exponents, and statistics of recurrence times. The rate of decay of correlations and the isolated spectrum of transfer operators in particular are notoriously difficult to estimate via iterative techniques.

2 Objects and behaviour of interest

In this section we give a summary of the definitions and properties of the objects we wish to estimate. Additional applications are outlined in §5.

2.1 Invariant measures

The fundamental object for ergodic-theoretic concepts is an *ergodic invariant probability measure* μ. One should think of $\mu(B)$ as representing the proportion of "mass" that is contained in the subset $B \subset M$. The *invariance condition* of μ is $\mu = \mu \circ T^{-1}$. This is a generalised "conservation of mass" equality; the mass distribution described by μ is preserved under the action of T, in the same way that area preserving maps preserve Lebesgue measure. *Ergodicity* is the measure-theoretic equivalent of topological transitivity. If two sets A and B both have positive μ-measure, then there is an $N \geq 0$ such that $\mu(A \cap T^{-N}B) > 0$; in other words, any region of positive μ-mass may be evolved forward to intersect any other region of positive μ-mass, with the intersection also having positive mass. Ergodic systems are indecomposable in the sense that μ-almost all starting points produce orbits with the same asymptotic distribution. The Birkhoff theorem [4, 52] tells us that the frequency with which orbits of μ-almost all starting points visit a set B is equal to the μ-measure of B; formally,

$$\lim_{N\to\infty} (1/N)\#\{0 \leq k \leq N-1 : T^k x \in B\} = \mu(B) \tag{4}$$

for μ-almost all $x \in M$. However, as in the case of dissipative systems, there may be a thin invariant attracting set Λ with $\mu(\Lambda) = 1$, but $m(\Lambda) = 0$. Equation (4) gives us no information about orbits starting off this invariant set.

Definition 2.1: An ergodic invariant probability measure is called a *natural* or *physical measure*, if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} f(T^k x) \to \int_M f \, d\mu \tag{5}$$

for all continuous $f : M \to \mathbb{R}$ and *Lebesgue-almost all* $x \in M$.

An alternative way of phrasing the definition of a physical measure is to state that the measure $\mu_N(x)$ of §1.3 converges weakly to μ as $N \to \infty$ for *Lebesgue almost all* $x \in M$. When talking about physical measures, equation (4) may be strengthened to:

Corollary 2.2:

$$\lim_{N\to\infty} \frac{1}{N}\#\{0 \leq k \leq N-1 : T^k x \in B\} = \mu(B) \tag{6}$$

for any subset $B \subset M$ *with* $\mu(\partial B) = 0$, *and for* Lebesgue almost all $x \in M$.

Corollary 2.2 says that "randomly chosen" initial conditions will with probability one, produce trajectories that distribute themselves according to the natural measure μ. Deterministic chaotic systems typically have infinitely many ergodic invariant measures (for example, a convex combination of δ measures on a periodic orbit), but only one physical measure.

Example 2.3: The linear mapping of the torus $T : \mathbb{T}^2 \circlearrowleft$ defined by $T(x,y) = (2x + y, x + y) \pmod 1$ has infinitely many ergodic invariant measures. For example, $\delta_{(0,0)}$, and $m =$ Lebesgue measure. Of all the ergodic invariant probability measures, only m satisfies (5).

2.2 Invariant sets

We approach invariant sets in a rather roundabout way, as we will describe them as the support of the physical measure, $\operatorname{supp}\mu$, where $\operatorname{supp}\mu$ is the smallest closed set having μ measure of 1 (equivalently, $x \in \operatorname{supp}\mu$ iff every open neighbourhood of x has positive μ measure). It is easy to show that $T(\operatorname{supp}\mu) = \operatorname{supp}\mu$ and $\operatorname{supp}\mu \subset T^{-1}(\operatorname{supp}\mu)$.

The reason for choosing $\operatorname{supp}\mu$ as the distinguished invariant set, rather than some invariant set defined via topological conditions is as follows. Let x be an arbitrary point in $\operatorname{supp}\mu$, and $B_\epsilon(x)$ be an open neighbourhood of size ϵ about x. By (5) orbits of *Lebesgue almost all initial points* visit $B_\epsilon(x)$ infinitely often and with a positive frequency of visitation. Meanwhile regions away from $\operatorname{supp}\mu$ are visited with frequency 0. As we wish to find the invariant set which appears on computer simulations of trajectories, it makes sense to consider only the support of μ and not some larger topologically invariant set.

2.3 Decay of correlations

The measure-theoretic analogy of topological mixing is (measure-theoretic) mixing. We say that a system (T,μ) is *mixing* if $\mu(A \cap T^{-N}B) \to \mu(A)\mu(B)$ as $N \to \infty$ for any pair of measurable subsets $A, B \subset M$. This condition says that the probability that a point x lies in a set B at time $t = 0$ and then moves to a set A at time $t = N$ (for large N) is roughly the product of the measures of the sets A and B. That is, for large N, the two events $\{x \in B\}$ and $\{T^N x \in A\}$ become statistically independent, or decorrelated. For dynamical systems with some degree of smoothness, this loss of dependence is often studied via the formula:

$$\mathcal{C}_{f,g}(N) := \left| \int f \circ T^N \cdot g \, d\mu - \int f \, d\mu \cdot \int g \, d\mu \right|, \tag{7}$$

where $f \in L^\infty(M,m)$ and $g : M \to \mathbb{R}$ has some smoothness properties. If one thinks of the functions $f, g : M \to \mathbb{R}$ as "physical observables" (output functions giving numerical information on some physical quantities of the system), then $C_{f,g}(N)$ quantifies the correlation between observing g at time t and f at time $t + N$. If $f = g$, we obtain what is commonly known as the *autocorrelation function*. For many chaotic systems, it is observed that $\mathcal{C}_{f,g}(N) \to 0$ at a geometric rate, and it is of interest to estimate this rate. The rate of decay can be interpreted variously as providing information on how quickly the system settles into statistically regular behaviour, how quickly transient behaviour disappears, and how quickly physical observables become decorrelated. For all of these interpretations, the physical measure μ is central as it provides the reference measure that describes statistical equilibrium for the system. Correlation decay has strong connections with *transfer operators* (or *Perron-Frobenius operators*) and the spectrum and corresponding eigenfunctions of these operators. We postpone further discussion of these objects until §3.3.

2.4 Lyapunov exponents

The Lyapunov exponents of a (piecewise differentiable) dynamical system describe the average local rates of expansion and contraction in phase space that are felt along typical trajectories. We may define them as the various values of λ that are observed[3] in the following expression, as v is varied over $\mathbb{R}^d$

$$\lambda := \lim_{N\to\infty} \frac{1}{N} \log \|D_x T^N(v)\| \tag{8}$$

As our dynamical system satisfies equation (5), it is immediate that Lebesgue almost all trajectories will produce the same d values of λ, and these values will be called *the*[4] Lyapunov exponents of T.

Later we will be discussing Lyapunov exponents of random matrix products. These can be defined in an analogous way.

2.5 Mean and variance of return times

The statistics of return times of orbits to subsets of phase space have not received a great deal of attention in the dynamical systems literature. A recent exception is [59], where regularity properties of return times are used to prove existence of physical measures and characterise rates of decay of correlations as exponential or algebraic. Suppose that one is given a subset $B \subset M$ with $\mu(B) > 0$. Every time a typical trajectory enters B, we note down the number of iterations since the orbit last entered B. In this way, we obtain a sequence of times $t_0, t_1, \ldots$ such that $T^{t_i}x \in B$, $i \geq 0$. We may now ask what is the mean and variance of this sequence of times. Formally, we define a function $R : M \to \mathbb{Z}^+$ by $R(x) = \inf\{k \geq 1 : T^k x \in B\}$, and use this return time function to define an induced map $\tilde{T}x = T^{R(x)}x$, where $\tilde{T} : B \circlearrowleft$. It is straightforward to show that $\mu|_B$ (defined by $\mu|_B(C) = \mu(C)/\mu(B)$ for $C \subset B$) is $\tilde{T}$-invariant. Therefore, we can define the mean return time

$$\mathbb{E}_{\mu|_B}(R) = \int_B R(x)\ d\mu|_B(x), \tag{9}$$

and the variance of the return times as

$$\mathrm{var}_{\mu|_B}(R) = \mathbb{E}_{\mu|_B}(R^2) - \mathbb{E}_{\mu|_B}(R)^2. \tag{10}$$

By Kac's theorem [34, 52], $\mathbb{E}_{\mu|_B}(R) = 1/\mu(B)$ provided μ is ergodic. A corresponding simple formula for the variance is not known.

3 Deterministic systems

In the first of two parts, we focus on deterministic systems. We show how each of the objects outlined in §2 will be approximated using the Markov model. We begin each subsection by simply outlining the constructions and the computations one performs in practice. At the end of each subsection, we state situations where rigorous results are obtained. While the Markov systems appear to retain many of the salient features of the original system, the smooth dynamics is completely lost and so it is not at all straightforward to prove strong approximation results.

[3] For a d-dimensional system, λ may only take on at most d different values.

[4] Lyapunov exponents depend very much on a reference measure (in this case the physical measure) in order to be defined.

3.1 Basic constructions

The region M is covered with a collection of connected sets $\mathfrak{P}_n = \{A_{n,1}, \ldots, A_{n,n}\}$ with nonempty interiors. This covering has the properties that (i) $\bigcup_{i=1}^n A_{n,i} = M$ and (ii) $\mathrm{Int}\, A_{n,i} \cap \mathrm{Int}\, A_{n,j} = \emptyset$ for $i \neq j$, where $\mathrm{Int}\, A$ denotes the interior of A. We think of each A_i as being a closed set, so that we do not have a partition in the strict sense, as the sets in the covering share boundaries.

We construct the matrix[5]

$$P_{n,ij} = \frac{m(A_{n,i} \cap T^{-1} A_{n,j})}{m(A_{n,i})}$$

as in (3).

Clearly, the quality of our Markov model depends heavily on the choice of partition. And it stands to reason that finer partitions produce better models (this will soon be made more precise). We will frequently produce a sequence of Markov models, each constructed from successively finer partitions, until we are satisfied with the accuracy of the estimates produced by the model. Sometimes one can produce better models by clever refinement strategies; these are discussed in §6.2.

3.2 Invariant measures and invariant sets

ROUGH IDEA: *The invariant density of the Markov chain approximates the physical invariant measure of T.*
REQUIRED COMPUTATION: *Calculate the fixed left eigenvector of P.*

We fix our partition $\{A_{n,1}, \ldots, A_{n,n}\}$ and calculate P_n and p_n, normalising p_n so that $\sum_{i=1}^n p_{n,i} = 1$. The value $p_{n,i}$ is the weight given to state i by the stationary distribution of the Markov chain. Since state i represents the set $A_{n,i}$ in our smooth space, we define an approximate invariant measure μ_n by assigning $\mu_n(A_{n,i}) = p_{n,i}$. Within the set $A_{n,i}$, we distribute the mass in any way we like. A common method is simply to spread the weight $p_{n,i}$ uniformly over $A_{n,i}$, so that formally,

$$\mu_n(B) := \sum_{i=1}^{n} \mu_n(A_{n,i}) \frac{m(A_{n,i} \cap B)}{m(A_{n,i})} \tag{11}$$

As we increase the number of partition sets n through refinement, some of these refined sets will be given zero measure by μ_n as an indication that trajectories spend no time or almost no time in these regions. Thus we expect that $\mathrm{supp}\, \mu_{n'} \subset \mathrm{supp}\, \mu_n$ for $n' > n$ (this can be made rigorous if T is a topological attractor [46]).

Further, since M is compact, the space of Borel probability measures on M (denoted $\mathcal{M}(M)$) is compact with respect to the weak topology. We may therefore continue this refinement procedure "forever" and extract a limiting measure μ^* as

$$\mu^* = \lim_{n \to \infty} \mu_n, \tag{12}$$

taking a convergent subsequence if necessary to define the limit. It is always assumed that $\max_{1 \leq i \leq n} \mathrm{diam}\, A_{n,i} \to 0$.

We have the following results concerning these approximations.

[5]This construction is often called Ulam's method as it was first proposed in [56] to use the matrix P_n to approximate invariant densities of interval maps.

Theorem 3.1: *Suppose that $T : M \circlearrowleft$ is continuous and has a physical measure μ. Let μ_n denote the approximate invariant measures and μ^* a weak limit of this sequence. Denote by S the intersection $\bigcap_{n=n_0}^{\infty} \operatorname{supp} \mu_n$. Then*

(i) μ^* *is T-invariant.*

(ii) $\operatorname{supp} \mu^* \subset \operatorname{supp} \mu_n$ *for all $n \geq 0$, and therefore* $\operatorname{supp} \mu^* \subset S$.

(iii) $\operatorname{supp} \mu \subset \operatorname{supp} \mu_n$ *for all $n \geq 0$, and therefore* $\operatorname{supp} \mu \subset S$.

PROOF:

(i) By noticing that the Markov models are small random perturbations [39, 41] of T, it is relatively simple to prove that μ^* is T-invariant [26, 14].

(ii) This is a straightforward consequence of weak convergence.

(iii) This is proven in [45, 46].

□

The first result of Theorem 3.1 says that weak limits of the sequence of numerically computed measures μ_n are T-invariant. In other words, we are in fact approximating *some* T-invariant probability measure. However, there is the question of just which invariant measure this is, as μ^* may be distinct from μ, the physical measure we wish to approximate. In this very general formulation, it is (as yet[6]) not possible to say that $\mu^* = \mu$.

Parts (ii) and (iii) of Theorem 3.1 say that at least the supports of the computed measures μ_n do contain the supports of both μ^* and μ, so that by stopping our computation at some finite n, we are not "losing" regions that are given positive measure by the physical measure μ.

Example 3.2 (The stiletto map): We introduce the Stiletto map [54] $T : \mathbb{R}^2 \circlearrowleft$.

$$T(x, y) = ((x + 1/3)\exp(-3x + 2) - 1/3 + y, 3x/10) \tag{13}$$

This map seems to possess chaotic dynamics on a fractal attracting set; see Figure 1. By selecting $M \subset \mathbb{R}^2$ to be a sufficiently small neighbourhood of the observed attracting set, it numerically appears that Lebesgue almost all $x \in M$ exhibit the invariant distribution described by the density of points in Figure 1. It is assumed that this distribution of points (in the infinite limit) describes the physical measure μ. We construct a Markov model using 1215 partition sets, where the sets are rectangles of equal size. The support of the resulting approximate invariant measure is shown in Figure 2, and the approximation itself is shown in Figure 3. We have used a relatively low number of partition sets for ease of viewing. Even for this crude model, there is good agreement between the distributions in Figures 1 and 3.

[6] A sign which may be taken as promise, or simply a state of ignorance, is that the author does not know of a continuous dynamical system T with physical measure μ for which $\mu^* \neq \mu$ (using reasonably "regular" partitions and refinements which keep the partition sets as approximate d-dimensional "cubes" of approximately the same shape and size).

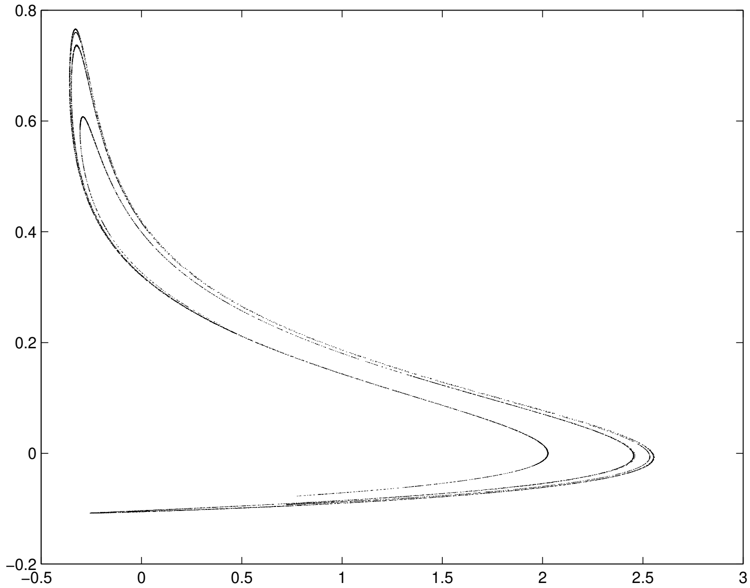

Figure 1: Trajectory of length 10^4 generated by the Stiletto map.

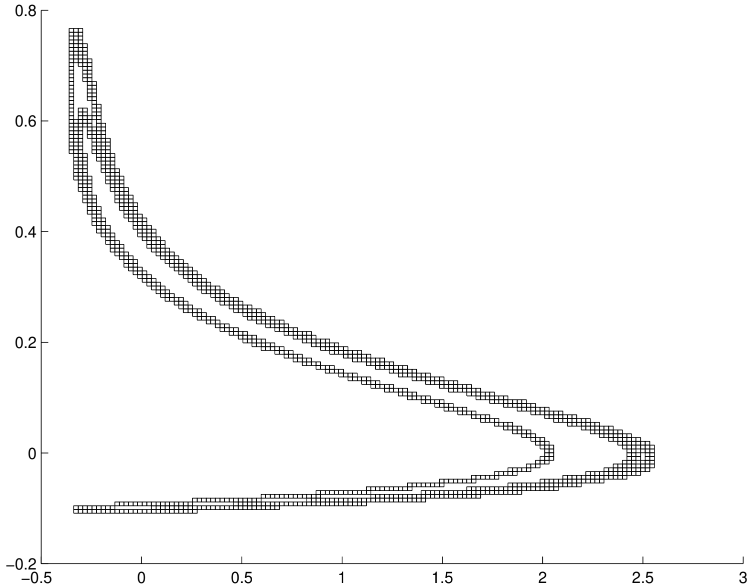

Figure 2: The support of an approximate invariant measure ($\operatorname{supp}\mu_{1215}$) shown as boxes.

3.2.1 Rigorous results

To state that $\mu^* = \mu$, we require further restrictions on the map T and the partition $\mathfrak{P}_n$; in general, even the question of *existence* of a physical measure for a given nonlinear system is still open. In one dimension, there is the classical existence result of Lasota and Yorke [43] that states that physical measures (probability measures with bounded densities) exist for piecewise C^2 interval maps $T : [0,1] \circlearrowleft$ with $\inf_x |T'(x)| > 1$. Li [44] first proved that under the additional constraint that $\inf_x |T'(x)| > 2$, these invariant densities could be approximated using Ulam's method in the sense that $\|\mu - \mu_n\|_{L_1} \to 0$ as $n \to \infty$. Since the publication of [44], there have been many variations of this basic result. In the setting

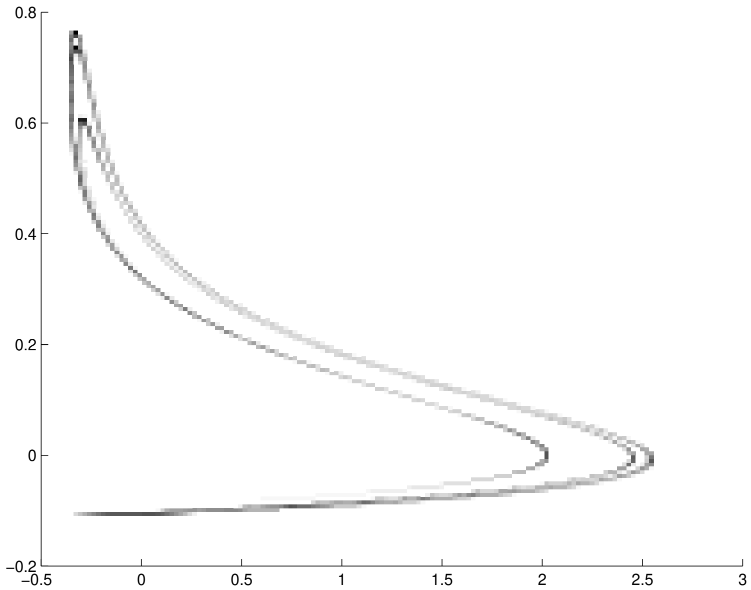

Figure 3: Representation of an approximate invariant (physical) measure (μ_{1215}) for the Stiletto map. Darker regions indicate higher density and more mass.

of [44], Keller [36] proved $\|\mu - \mu_n\|_{L^1} = O(\log n/n)$. Recent results (often under additional "ontoness" assumptions) have focussed on explicit error bounds for the difference $\|\mu - \mu_n\|_{L^1}$; [35, 49, 17].

For higher dimensional uniformly expanding systems, very roughly speaking, the paper of Boyarsky and Gora [27] and Ding [12] mirror those of [43] and [44]. There are several additional technical constraints on the map T and the partitions $\mathfrak{P}_n$ that we do not discuss. Again under some ontoness conditions, Murray [48] applies the methods of [49] to provide error bounds for $\|\mu - \mu_n\|_{L^1}$.

For uniformly hyperbolic systems, the author [14, 18] shows that $\mu_n \to \mu$ weakly (resp. in L^1) when the physical measure μ is singular (resp. absolutely continuous), provided that the partitions $\mathfrak{P}_n$ are Markov partitions.

A combination of theory and numerics [16] suggests that the convergence rate $O(\log n/n)$ holds in reasonable generality for systems with good mixing properties.

3.3 Decay of correlations and spectral approximation

Rough Idea: *The spectrum of the matrix P approximates the spectrum of the Perron-Frobenius operator.*
Required Computation: *Calculate the eigenvalues of P.*

We begin by noting that we have an alternative formulation of (7) in terms of the *Perron-Frobenius operator*[7] $\mathcal{P} : L^1(M, m) \circlearrowleft$.

Lemma 3.3: *Let $\mathcal{F}$ be a class of real-valued functions preserved*[8] *by $\mathcal{P}$. Let $\sigma(\mathcal{P})$ denote the spectrum of $\mathcal{P}$ when considered as an operator on $\mathcal{F}$, and set $r = \sup\{|z| : z \in \sigma(\mathcal{P}) \setminus \{1\}\}$. Then there is a constant $C < \infty$ such that $\mathcal{C}_{f,g}(N) \le Cr^N$ if $g \in \mathcal{F}$ and $f \in L^\infty$.*

[7] See [42, 6] for definitions and properties of the Perron-Frobenius operator.

[8] For example, if T is C^γ, then $C^{\gamma-1}(M, \mathbb{R})$ is preserved by $\mathcal{P}$, and if $T : [0, 1] \circlearrowleft$ is a *Lasota-Yorke map*, then functions of bounded variation are preserved by $\mathcal{P}$.

This result says that we may bound the rate of decay of correlations by the maximal non-unit spectral value for the operator $\mathcal{P}$; see Figure 4 (upper right) for the typical spectral plot we have in mind. The Ulam matrix P_n may be thought of as a projection of $\mathcal{P}$ onto a finite dimensional space. Naively, then, we may think that the spectrum of the matrix P_n will approximate the spectrum of the Perron-Frobenius operator. This would be very useful, as it is simple to compute the spectrum of P_n since the matrix is very sparse and there are numerical routines to compute only eigenvalues which are large in magnitude (these are the ones that are principally of interest). Furthermore, the eigenfunctions of $\mathcal{P}$ corresponding to large eigenvalues are also of interest, as they indicate the sorts of mass distributions that approach the equilibrium distribution (the physical measure) at the slowest possible rate. Perhaps we can also approximate these slowly mixing distributions (eigenfunctions of $\mathcal{P}$) as eigenvectors of the matrix P. If we can, there is the question of what these slowly mixing distributions represent. One generally thinks that the rate of mixing is determined by expansion properties of the dynamical system; that is, the more expansive (or more "chaotic"), the greater the rate of decay. But the existence of distributions which mix at a rate *slower* than that dictated by the minimal expansion rate of the system, presents a seemingly paradoxical situation. Arguments in [7, 11, 9] suggest that these distributions describe "macroscopic structures" embedded within the dynamics, which exchange mass very slowly and work against the chaoticity.

3.3.1 Rigorous results

There are two situations where these ideas can be made rigorous. First, for one-dimensional maps, there is the recent result of Keller and Liverani [37].

Theorem 3.4: *Let BV denote the space of functions of bounded variation on $[0,1]$. Suppose $T : [0,1]\circlearrowleft$ is an expanding, piecewise C^2 interval map, with $\inf_{x\in[0,1]} |T'(x)| = \alpha > 1$. Then isolated eigenvalues of $\mathcal{P} : BV\circlearrowleft$ outside the disk $\{|z| \leq 1/\alpha\}$ and the corresponding eigenfunctions are approximated by eigenvalues and the corresponding eigenvectors of P_n (eigenfunction convergence in the L^1 sense). The convergence rate for the eigenvectors to the eigenfunctions for eigenvalues $z \in (\alpha, 1)$ is $O(n^{-r})$, where $0 < r(z) < 1$, while the eigenvector approximating the invariant density ($z = 1$) converges like $O(\log n/n)$.*

Example 3.5 (The double wigwam map): We introduce the map $T : [0,1]\circlearrowleft$ defined by

$$T(x) = \begin{cases} -2x + 1 - \sin(4\pi x)/30, & 0 \leq x < 1/4 \\ 3(x - 1/4) + 1/4 + \sin(4\pi(x - 1/4))/10, & 1/4 \leq x < 1/2 \\ 3(x - 1/2) - \sin(4\pi(x - 1/2))/10, & 1/2 \leq x < 3/4 \\ -2(x - 1) - \sin(-4\pi)/30, & 3/4 \leq x \leq 1 \end{cases},$$

the graph of which is shown in the upper left frame of Figure 4. Theorem 3.4 will be used to show that the spectrum of $\mathcal{P} : BV\circlearrowleft$ contains a non-trivial isolated eigenvalue, and therefore a rate of decay *slower* than that prescribed by the minimal expansion rate. As T is a *Lasota-Yorke* map, the classical result of [43] tells us that T possesses an invariant density (this is the density of the physical measure). The result of Li [44] tells us that this invariant density may be approximated by eigenvectors of the Ulam matrices. The bottom left frame of Figure 4 shows a plot of an approximation of the invariant density using an equipartition of $[0,1]$ into 512 sets. The upper right frame shows the spectrum of the resulting 512×512 matrix. The large dotted circle denotes $\{|z| = 1\}$, and the dash-dot inner circle shows the upper bound for the essential spectral radius for this map $\{|z| = 0.6325\}$, $0.6325 = 1/\inf_{x\in[0,1]}|T'(x)|$. The cross shows an eigenvalue of P_{512} that is clearly outside this inner region and therefore corresponds to a true isolated eigenvalue of $\mathcal{P}$. The eigenfunction for this isolated eigenvalue is plotted in the lower right frame.

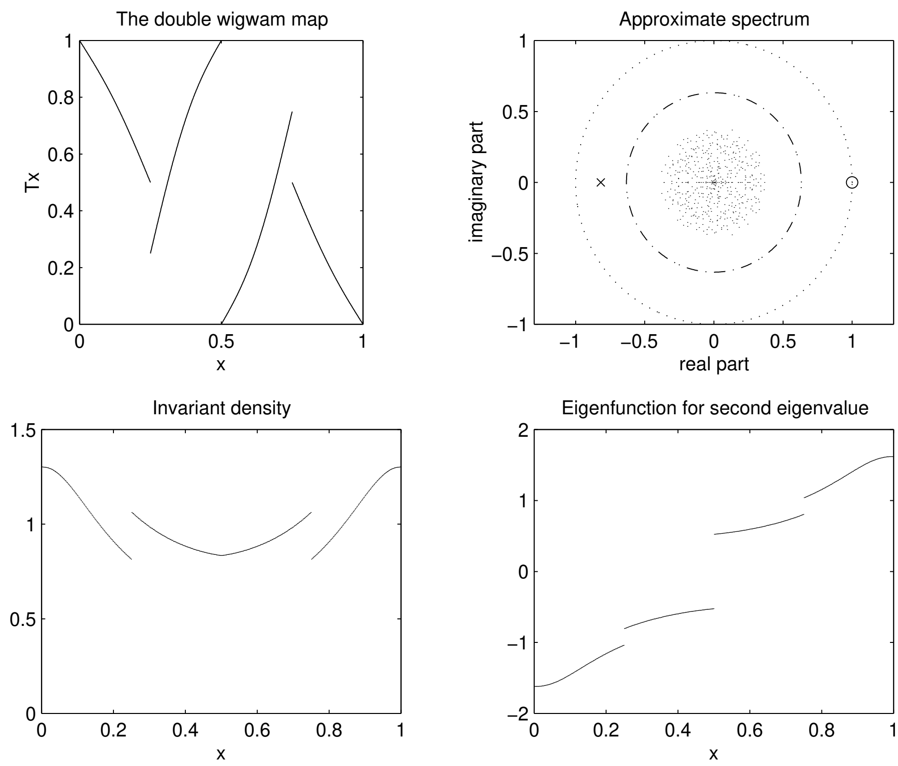

Figure 4: (upper left): Graph of T; (upper right): Spectrum of 512×512 transition matrix P_{512}, the small circle represents the eigenvalue 1, the small cross represents another isolated eigenvalue; (lower left): Plot of the invariant density of T (the eigenfunction for the eigenvalue 1); (lower right): Plot of the eigenfunction for the second isolated eigenvalue.

What about higher-dimensional systems? For uniformly hyperbolic systems, a standard technique is to factor out the stable directions and consider only the action of T along unstable manifolds W^u. This induces an expanding map $T_E : W^u \circlearrowleft$ with corresponding Perron-Frobenius operator $\mathcal{P}_E$. We have the following result.

Theorem 3.6 ([15]): *Let $T : M \circlearrowleft$ be $C^{\gamma+1}$, $0 < \gamma \leq 1$, uniformly hyperbolic, and possess a nice[9] Markov partition. Construct P_n by setting $\mathfrak{P}_n$ to be a refinement of this Markov partition, and consider $\mathcal{P}_E$ to act on the function space $C^\gamma(W^u, \mathbb{R})$. Isolated eigenvalues of $\mathcal{P}_E : C^\gamma(W^u, \mathbb{R}) \circlearrowleft$ and the corresponding eigenfunctions are approximated by eigenvalues and the corresponding eigenvectors of P_n (eigenfunction convergence in the smooth C^γ sense). The rate of convergence of both the eigenvalues and the eigenvectors to the isolated eigenvalues and corresponding eigenfunctions of $\mathcal{P}$ is $O(1/n^r)$, where $0 < r < 1$ depends[10] only on maximal and minimal values of the derivative of T in unstable directions, and convergence of the eigenfunctions (including the invariant density) is with respect to the stronger smooth norm.*

[9]See [18] for a definition of nice.

[10]Let $1/\ell$ (resp. $1/L$) denote the minimal (resp. maximal) stretching rate of T in unstable directions. Then $r = \log(\ell)/2\log(L)$ and convergence is in the $\|\cdot\|_{\gamma/2}$ norm.

In the hyperbolic case, a bound for the rate of decay for the full map T may be extracted from the rate of decay for the induced expanding map T_E (see [15]). The case of uniformly expanding T is a simple special case of Theorem 3.6.

3.4 Lyapunov exponents and entropy

Rough Idea #1: *Lyapunov exponents may be calculated by averaging local rates of expansion according to the physical measure (a spatial average).*
Rough Idea #2: *The Lyapunov exponents of the Markov model approximate the Lyapunov exponents of T.*
Rough Idea #3: *The local stretching rates and dynamics may be encoded in the matrix P to provide Lyapunov exponent and entropy estimates.*

We begin by recalling that for one-dimensional systems, the expression (8) may be rewritten as

$$\lambda = \int_M \log |T'(x)| \, d\mu(x) \tag{14}$$

by a straightforward application of (5) with $f(x) = \log |T'(x)|$. Thus, once we have an estimate of the physical measure μ, it is easy to compute an approximation of λ via

$$\lambda_n := \int_M \log |T'(x)| \, d\mu_n(x) \approx \sum_{i=1}^{n} \log |T'(x_{n,i})| \cdot p_{n,i}, \tag{15}$$

where $x_{n,i} \in A_{n,i}$ (for example, $x_{n,i}$ could be the midpoint of the interval $A_{n,i}$). The error bounds for $\|\mu - \mu_n\|_1$ (when available) immediately translate into rigorous bounds for the error $|\lambda - \lambda_n|$. We now turn to multidimensional systems.

3.4.1 Approach #1

The direct use of the physical measure for Lyapunov exponent computation may be extended to higher dimensional systems, by rewriting (8) as

$$\lambda = \int_M \log \|D_x T(w_x)\| \, d\mu(x) \tag{16}$$

where $\{w_x\}_{x\in M}$ is a family of unit vectors in $\mathbb{R}^d$ satisfying the identity $D_xT(w_x) = w_{Tx}$; different families yield the different Lyapunov exponents (see [53, 21, 22] for details). For the remainder of this section, we consider the problem of finding the largest Lyapunov exponent λ^1; the remaining exponents may then be found via standard methods involving exterior products. We denote the vector field corresponding to λ^1 by $\{w_x^1\}$. The vector w_x^1 is the eigenvector of the limiting matrix $\Lambda_x := \lim_{N\to\infty}((D_xT^{-N})^\top(D_xT^{-N}))^{1/2N}$ corresponding to the smallest eigenvalue (in magnitude) [53]. One may approximate the vector w_x^1 by computing the smallest eigenvector of the matrix $\Lambda_{N,x} := (D_xT^{-N})^\top(D_xT^{-N})$ for some small finite N. For $N = 7$, the approximate vector field $\{w_{N,x}^1\}$ for the Stiletto map is shown in Figure 5. Thus we may compute an approximation to (16) by

$$\lambda_{n,N}^1 = \sum_{i=1}^{n} \log \|D_{x_{n,i}}T(w_{N,x_{n,i}}^1)\| \cdot p_{n,i} \tag{17}$$

where $w_{N,x}^1$ denotes the eigenvector obtained from $\Lambda_{N,x}$; see Table 1.

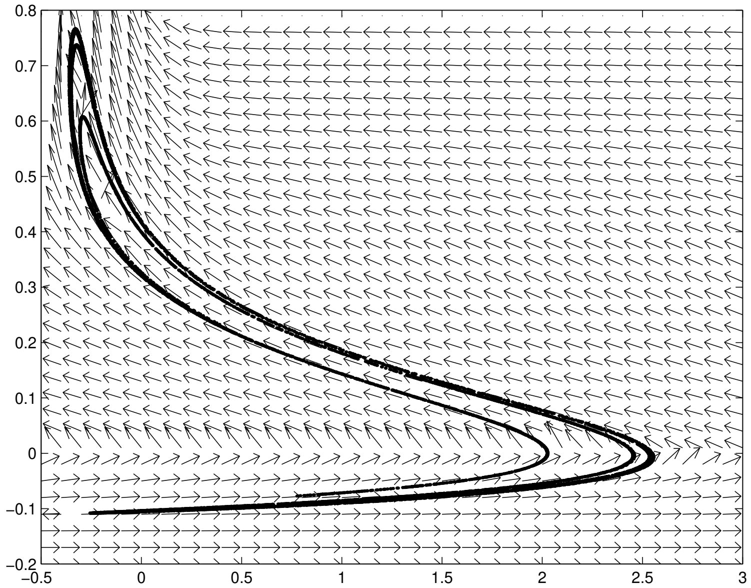

Figure 5: Approximate vector field of the "Lyapunov directions" $\{w_x^1\}$ corresponding to the largest Lyapunov exponent. In this two-dimensional example, w_x^1 should be tangent to the unstable manifold passing through x.

Table 1: Lyapunov exponent estimates for the Stiletto map using the method of [22] and the approximate invariant measure of Figure 3

N	1	2	3	4	5	6	7	8
$\lambda^1_{1215,N}$	0.6826	0.4107	0.4458	0.4536	0.4530	0.4528	0.4529	0.4528

A recent method put forward by Aston and Dellnitz [2] notes that

$$\lambda^1 = \inf_{N\geq 1} \frac{1}{N}\int_M \log\|D_x T^N\|\; d\mu(x). \tag{18}$$

They therefore propose the approximation

$$\lambda^1_{n,N} := \sum_{i=1}^{n} \log\|D_{x_{n,i}} T^N\| \cdot p_{n,i} \tag{19}$$

Table 2: Lyapunov exponent estimates for the Stiletto map using the method of [2]

N	2	4	8	16	32	64	128	256
$\lambda^1_{1215,N}$	0.6303	0.4530	0.4476	0.4322	0.4435	0.4391	0.4436	0.4441

In practice, $\lambda^1_{n,N}$ greatly overestimates λ^1, and so to speed up convergence, one defines $\lambda^{1'}_{n,M} := 2\lambda_{n,2^M} - \lambda_{n,2^{M-1}}$, $M \geq 1$; the values $\lambda^{1'}_{n,M}$ are observed to have better convergence properties; see Table 2.

3.4.2 Approach #2

Through our Markov modelling process, we have approximated the dynamics of T as a large Markov chain governed by P_n. To each state i in the Markov chain, we may associate the Jacobian matrix $D_{x_{n,i}}T$, where $x_{n,i}$ denotes the centre point (for example) of the partition set $A_{n,i}$. We now consider the Lyapunov exponents of this Markov chain; as we move from state to state along a random orbit of the chain, we multiply together the matrices we have assigned to these states. This produces a random composition of matrices, and the theory of Lyapunov exponents for random matrix products is well developed (see §3.4 [1], for example). Continuing our theme of "the deterministic dynamics is well approximated by the Markov model", we compute the top Lyapunov exponent of the Markov model and use this as an approximation of λ^1. The top Lyapunov exponent for this Markov chain is given by an equation of the form:

$$\lambda_n^1 := \sum_{i=1}^{n} \left(\int_{\mathbb{RP}^{d-1}} \log \|D_{x_{n,i}}T(v)\| \; d\xi_{n,i}^1(v) \right) \cdot p_{n,i} \tag{20}$$

where $\xi_{n,i}^1$ is a probability measure on $\mathbb{RP}^{d-1}$ ($d-1$-dimensional real projective space, or "the space of directions in $\mathbb{R}^d$"); see [21, 20] for details.

Table 3: Lyapunov exponent estimates for the Stiletto map using the method of [21]

"Resolution of ξ^1"	10	20	30	40	60	80	100	200
$\lambda^1_{1215,N}$	0.4322	0.4001	0.4260	0.4419	0.4637	0.4408	0.4458	0.4456

Whereas the vector $w^1_{x_{n,i}}$ indicates a single direction in which to measure the stretching caused by $D_{x_{n,i}}T$, the measure $\xi^1_{n,i}$ indicates a distribution of directions in which to measure the stretching. This distribution is essential, as the vectors w^1_x often vary within partition sets (for example, near the "toe" area of the Stiletto attractor where the unstable manifold bends sharply), and so it is necessary to "average" these directions *within* partition sets, rather than take a single direction as in Approach #1. Roughly speaking, the distribution $\xi^1_{n,i}$ can be thought of as a histogram of the vectors w^1_x, $x \in A_{n,i}$. For reasons of space, we refer the reader to [21, 20], in which the details of the calculation of λ^1_n are spelt out.

3.4.3 Rigorous results (Approach #3)

The two above approaches are not rigorous. Approach #1 is not rigorous because we do not know that $\mu_n \to \mu$ (although numerically this appears to happen). Approach #2 additionally suffers from the possible sensitivity of the Lyapunov exponents to perturbations of the system – for example, the perturbation we used to create the Markov model – however, this sensitivity is also rarely observed numerically. In the uniformly expanding or uniformly hyperbolic case, if we use a Markov partition to construct our transition matrix, one can prove convergence of Lyapunov exponent estimates to the true value, and additionally, obtain rigorous estimates of the metric entropy and escape rate / pressure of the system.

Theorem 3.7 ([18]): *Construct $Q_n = m(A_{n,i} \cap T^{-1}A_{n,j})/m(A_{n,j})$ using a nice Markov partition. Let ϱ_n denote the largest eigenvalue of Q_n, and v_n the corresponding right eigenvector. Construct the stochastic matrix $P_{n,ij} = Q_{n,ij}v_{n,j}/\varrho_n v_{n,i}$, and compute the fixed left* *eigenvector p_n of P_n. Define μ_n as in (11) with $\mu_n(A_{n,i}) = p_{n,i}$. Define*

$$\lambda_n := -\sum_{i,j=1}^{n} p_{n,i} P_{n,ij} \log Q_{n,ij} \tag{21}$$

$$h_n := \log \varrho_n + \lambda_n \tag{22}$$

Then as $n \to \infty$, $\mu_n \to \mu$, $\lambda_n \to \sum_{\lambda^{(i)}>0} \lambda^{(i)}$ (the sum of the positive Lyapunov exponents), $h_n \to h_\mu(T)$ (the measure-theoretic entropy of T with respect to μ), and $\varrho_n \to P(T)$ (the topological pressure of T). Convergence rates are also available.

3.5 Mean and variance of return times

Rough Idea: *The mean and variance of return times calculated for the Markov model approximate those of T.*
Required Computations: *Calculate the fixed left eigenvector of P, and solve a linear equation of the form $Ax = b$.*

We have the following abstract result:

Theorem 3.8: *Let $T : X \circlearrowleft$ preserve an ergodic invariant measure μ. Let $B \subset X$, with $0 < \mu(B) < 1$, and set $B^c = X \setminus B$.*

(i)

$$\mathbb{E}_{\mu|_B}(R) = 1/\mu(B), \tag{23}$$

(ii)

$$\mathrm{var}_{\mu|_B}(R) = \frac{1-\mu(B)}{\mu(B)} \left(2\mathbb{E}_{\mu|_{B^c}}(R) - 1/\mu(B)\right) \tag{24}$$

Part (i) is due to Kac [34] and (ii) to Blum and Rosenblatt [5] (though we have used the version given in [13]; see also [58]). Theorem 3.8 reduces the problem of calculating $\mathrm{var}_{\mu|_B}(R)$ to a calculation of $\mathbb{E}_{\mu|_{B^c}}(R)$, the expected first absorption time into B for points in B^c. The calculation of first absorption times is simple to do when the dynamical system is a finite state Markov chain.

Let B be a subset of the phase space M. This set will be covered by a collection of partition sets $A_{n,i_1}, \ldots, A_{n,i_q}$; in practice, one obtains more accurate results if it can be arranged so that B is exactly a union of some collection of partition sets.

We now apply Theorem 3.8 to our Markov model[11], setting B to be the collection of states $\{i_1, \ldots, i_q\}$. Using (23) and (24) it is possible to calculate *exact* values for the mean and variance of the return times to the collection of states $\{i_1, \ldots, i_q\}$. Appealing then to our guiding principle that our Markov model approximates our original map T, we take the mean and variance of the return times to the states $\{i_1, \ldots, i_q\}$ as the approximate mean and variance for the return times to the set $B \subset M$.

We now outline the necessary computations for our Markov model. For simplicity, we assume that the states of our Markov chain have been reordered, so that the states $\{i_1, \ldots, i_q\}$ now have labels $\{n-q+1, \ldots, n\}$; for the remainder of this section, the matrix P_n will denote this reordered matrix. To calculate the mean of the recurrence time (denoted $\mathcal{M}_n$):

[11] Formally, we set $X = \Omega = \{(\omega_0, \omega_1, \ldots) : P_{\omega_i, \omega_{i+1}} > 0, \omega_i \in \{1, \ldots, n\}, i \geq 0\}$, $T = \sigma : \Omega \circlearrowleft$, the left shift on Ω, and $\mu = \mathbb{M}$, the Markov measure defined on cylinder sets by $\mathbb{M}([\omega_i, \ldots, \omega_{i+t}]) = p_{\omega_i} P_{\omega_i, \omega_{i+1}} \cdots P_{\omega_{i+t-1}, \omega_{i+t}}$. Then $B = [\omega_{i_1}] \cup \cdots \cup [\omega_{i_q}]$.

(i) Calculate the invariant density p_n for the Markov chain governed by P_n.

(ii) Set $\mathcal{M}_n := 1/\sum_{i=n-q+1}^{n} p_{n,i}$.

To calculate the expected absorption time and the variance of the recurrence time (denoted $\mathcal{A}_n$ and $\mathcal{V}_n$ respectively):

(i) Write P_n in the block form

$$P_n = \begin{pmatrix} Q_n & U_n \\ V_n & Y_n \end{pmatrix} \tag{25}$$

where the matrix Q_n is a $(n-q) \times (n-q)$ matrix giving the transition probabilities between states in our Markov model *not* corresponding to the set $B \subset M$.

(ii) Calculate the solution τ_n to the linear equation (I_{n-q} is the $(n-q) \times (n-q)$ identity matrix)

$$(I_{n-q} - Q_n)\tau_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}^\top, \tag{26}$$

(iii) Set $\mathcal{A}_n := (\sum_{i=1}^{n-q} p_{n,i}\tau_{n,i})/(\sum_{i=1}^{n-q} p_{n,i})$.

(iv) Set $\mathcal{V}_n := (\mathcal{M}_n - 1)(2\mathcal{A}_n - \mathcal{M}_n)$.

See [19] for further details.

The number $\tau_{n,i}$ is an approximation of the average time required for a point in $A_{n,i}$ to move into B, and is often of interest in itself.

Example 3.9 (The bouncing ball): We study a two-dimensional map of a cylinder $T : S^1 \times \mathbb{R} \circlearrowleft$ that describes the evolution of a ball bouncing on a sinusoidally forced table. We set

$$T(\phi, v) = (\phi + v, \alpha v - \gamma \cos(\phi + v)) \tag{27}$$

where $\phi \in [0, 2\pi)$ represents the phase of the table at impact, $v \in \mathbb{R}$ the velocity of the ball just after impact with the table, and T represents the evolution from one impact to the next; see [28] for details. We set $\alpha = 0.5$ and $\gamma = 10$ for the remainder of this example. Figure 6 shows a typical orbit of the system, and Figure 7 shows an approximation of the "physical" invariant measure μ; again, there is good agreement between the two distributions. We suppose that we are interested in the time between successive impacts where the velocity of the ball is very low; that is, we create a time series by counting the time intervals between instances when the ball leaves the table with a velocity of magnitude less than 1.25. Thus $B = S^1 \times [-1.25, 1.25]$ in the earlier notation. Performing the analysis described above, Table 4 shows the results for various partition refinements.

Compare these values with 9.830 ± 0.021 and 9.373 ± 0.032 the mean and variance respectively obtained directly from 10 orbits of length 10^6 (plus/minus one standard deviation of the 10 values obtained).

The result of the calculation of τ_n is shown in Figure 8. It is clear that there is a sharp divide between areas which return to low velocities relatively quickly (the very dark strips in Figure 8) and those areas that take longer to return. A histogram plot of τ reveals that if a point does not return very quickly to B, then it takes a much longer time.

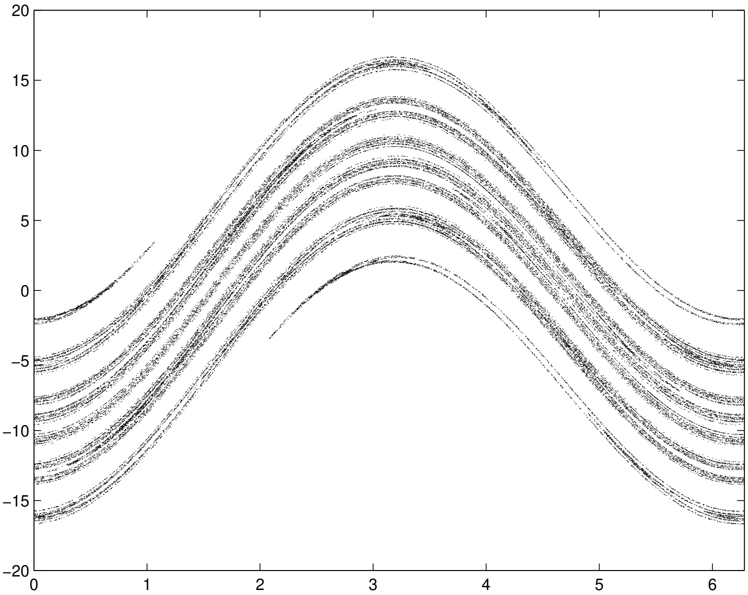

Figure 6: Plot of orbit of length 50000 for the bouncing ball map.

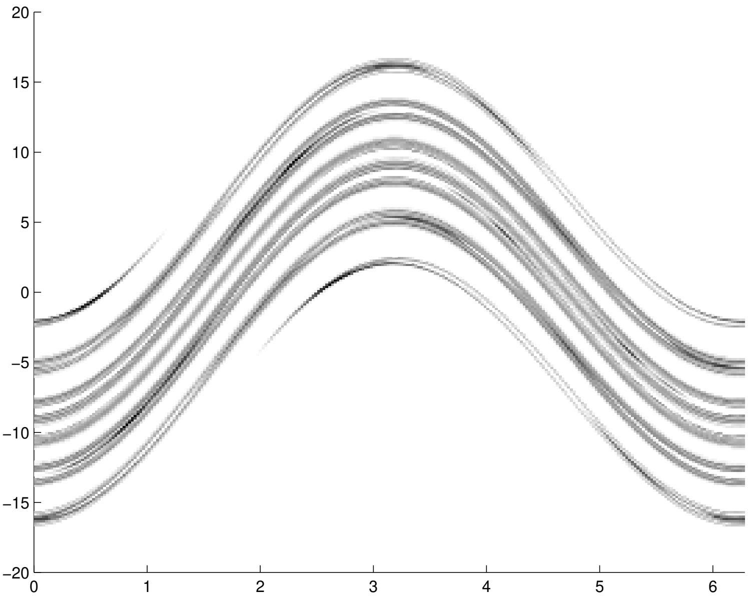

Figure 7: Approximate invariant measure for the bouncing ball map using 16566 partition sets. Darker regions indicate higher density.

Table 4: Estimates of the mean and variance of return times to the set $B = S^1 \times [-1.25, 1.25]$

Number of partition sets n	1654	5086	16566
Mean $\mathcal{M}_n$	10.33	9.97	9.85
Root Variance $\sqrt{\mathcal{V}_n}$	9.82	9.45	9.33

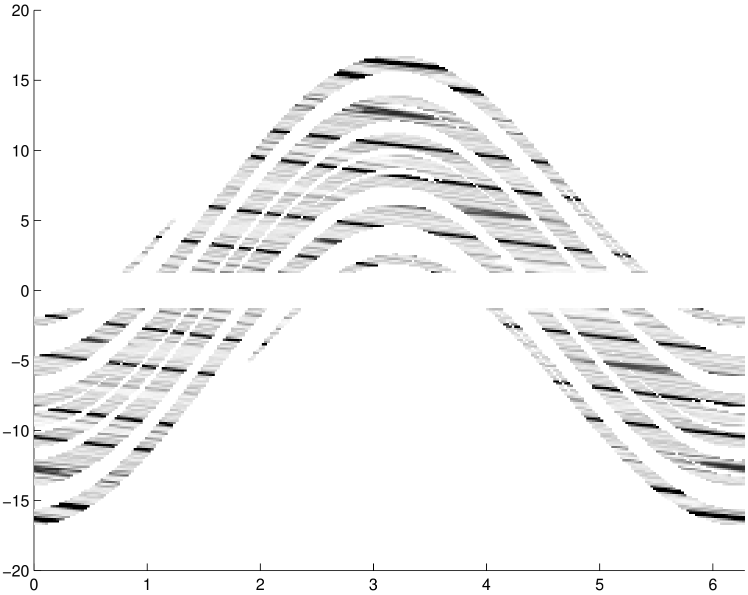

Figure 8: Approximate absorption times into the set $S^1 \times [-1.25, 1.25]$. Faster absorption is indicated by darker shading.

4 Random Systems

We now discuss extensions of our Markov modelling process to random dynamical systems. Instead of having a single deterministic mapping $T : M \circlearrowleft$, we now have a collection of mappings $\{T_1, \dots, T_r\}$, where $T_k : M \circlearrowleft$ for $k = 1, \dots, r$. A random orbit $\{x_N\}_{N=0}^{\infty}$ is defined by setting

$$x_N = x_N(k_{N-1}, \dots, k_0, x_0) := T_{k_{N-1}} \circ \cdots \circ T_{k_1} \circ T_{k_0} x_0, \quad N \geq 1, \tag{28}$$

where the indices $k_0, k_1, \dots \in \{1, \dots, r\}$ are generated by a stationary stochastic process. We will be considering two situations; namely where the indices are generated by iid processes and Markov processes.

The former case corresponds to the situation where at each time step, a map is selected at random (according to some fixed probability distribution on the numbers $\{1, \dots, r\}$), independently of any previously applied maps. We shall say that the probability of selecting the map T_k at any given time is $w_k \geq 0$; naturally $\sum_{k=1}^{r} w_k = 1$. The probability of the sequence $T_{k_{N-1}} \circ \cdots \circ T_{k_0}$ occurring is simply $w_{k_{N-1}} \cdots w_{k_0}$. By extending this product to infinite sequences of indices, we obtain a probability measure $\mathbb{P}$ on $\{1, \dots, r\}^{\mathbb{Z}^+}$. The monograph [40] lays down much of the basic theory concerning such dynamical systems, including generalisations of standard dynamical systems quantities to the random situation.

In the Markov case, the probability with which a map is selected at each time step depends only on the previously applied map[12]. We shall say that the probability of selecting a map T_l *given that the map* T_k *was selected at the last time step*, is $W_{kl} \geq 0$; we require that $\sum_{l=1}^{r} W_{kl} = 1$, and so W is a stochastic matrix governing the Markov chain that produces our random sequence of indices. The probability of the sequence $T_{k_{N-1}} \circ \cdots \circ T_{k_0}$ occurring is $w_{k_0} W_{k_0 k_1} \cdots W_{k_{N-2} k_{N-1}}$; by extending this to infinite sequences, we obtain a probability measure (a Markov measure) on $\{1, \dots, r\}^{\mathbb{Z}^+}$, which we also denote by $\mathbb{P}$.

[12]If one desires to treat Markov processes with longer memory, they may be written in terms of a first order Markov chain in the standard way.

Examples 4.1:

(i) Consider the bouncing ball map of the last section. Suppose that our ball is non-uniform, and that one side is more "springy" than the other. Sometimes, the ball will land on the springy side, and sometimes it will land on the not-so-springy side. Which side the ball lands on determines the value of α, and so at each time step there is a random choice of α, and therefore an application of either $T_{\alpha_{\mathrm{springy}}}$ or $T_{\alpha_{\mathrm{not-so-springy}}}$. We return to this example later.

(ii) A set of maps $\{T_1, \ldots, T_r\}$ could arise as perturbations of a single map T via $T_k x := Tx + \epsilon_k$, where $\epsilon_k \in \mathbb{R}^d$ is a perturbation. We choose a probability vector $(w_1, \ldots, w_r)$ where the value w_k represents the probability of our map T encountering the perturbation ϵ_k. A random iid composition of the $\{T_k\}$ models a deterministic system subjected to small iid perturbations.

(iii) Random dynamical systems can also arise in the context of dynamical systems with inputs. The effect of an input is essentially to produce different dynamics (in other words, a different map T_k) at the time step in which it occurs. If the model is truly random, these inputs could occur according to an iid process or Markov process. However, more structured sets of inputs can also be modelled by Markov processes, for example, where a randomly selected input triggers a fixed sequence of inputs before another random input is selected.

We now define what is meant by an invariant measure for our random system.

Definition 4.2: Let $\Omega = \{1, \ldots, r\}^{\mathbb{Z}^+}$, and for $\omega = (\omega_0, \omega_1, \omega_2, \ldots) \in \Omega$, define the left shift $\sigma : \Omega \circlearrowleft$ by $(\sigma\omega)_i = \omega_{i+1}$. The probability measure $\mathbb{P}$ on Ω introduced earlier is σ-invariant. Define the skew product $\tau : \Omega \times M \circlearrowleft$ by $\tau(\omega, x) = (\sigma\omega, T_{\omega_0} x)$; our random orbits $\{x_N\}_{N \geq 0}$ may be written as $x_N = \mathrm{Proj}_M(\tau^N(\omega, x_0))$, where Proj_M denotes the canonical projection onto M.

We will say that a probability measure μ on M is an *invariant measure* for our random system, if there exists a τ-invariant probability measure $\tilde{\mu}$ on $\Omega \times M$ such that

(i) $\tilde{\mu}(E \times M) = \mathbb{P}(E)$ for all measurable $E \subset \Omega$, and

(ii) $\tilde{\mu}(\Omega \times B) = \mu(B)$ for all measurable $B \subset M$.

Definition 4.3: A probability measure μ is called a *natural* or *physical* measure for a random dynamical system if μ is defined as $\mu(B) = \tilde{\mu}(\Omega \times B)$ where $\tilde{\mu}$ is a τ-invariant probability measure satisfying

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N-1} f(\tau^k(\omega, x)) \to \int_{\Omega \times M} f(\omega, x) \, d\tilde{\mu}(\omega, x) \tag{29}$$

for all continuous $f : \Omega \times M \to \mathbb{R}$, and $\mathbb{P} \times m$ almost all $(\omega, x) \in \Omega \times M$.

Remark 4.4: If we choose the continuous test function f in (29) to be independent of ω, then we have the simple consequence that:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{j=0}^{N-1} f(T_{k_j} \circ \cdots \circ T_{k_0} x) \to \int_M f \, d\mu \tag{30}$$

for Lebesgue almost all $x \in M$ and $\mathbb{P}$ almost all random sequences of maps.

By setting $f(x) = \chi_A(x)$, where $A \subset M$ is such that $\mu(\partial A) = 0$, then

$$\frac{1}{N}\,\mathrm{card}\{0 \le j \le N-1 : T_{k_j} \circ \cdots T_{k_1} \circ T_{k_0} x \in A\} \to \mu(A),$$

again for Lebesgue almost all $x \in M$ and $\mathbb{P}$ almost all random sequences.

In rough terms, this says that if you plot the points in a random orbit defined by (28), then for Lebesgue almost all starting points x_0 and $\mathbb{P}$ almost all random sequences of maps, one obtains the same distribution of points. From the point of view of analysing the average behaviour of the random system, this is the correct distribution to approximate. The physical measure μ is usually *not* invariant under *any* of the individual transformations T_k, in the sense that $\mu \circ T_k^{-1} \neq \mu$. However, μ is "invariant on average", which by heavily abusing notation may be written as $\mathbb{E}(\mu \circ T_k^{-1}) = \mu$. In the iid case, this formula is entirely accurate, as there the invariance condition is simply $\sum_{k=1}^r w_k \mu \circ T_k^{-1} = \mu$.

4.1 Basic constructions

We must construct a transition matrix

$$P_n(k) = \frac{m(A_{n,i} \cap T_k^{-1} A_{n,j})}{m(A_{n,i})} \tag{31}$$

for each of the maps T_k.

Remark 4.5: An alternative definition of the matrix $P_n(k)$ is as follows. Within each set $A_{n,i}$ select a single point $a_{n,i}$. Then set

$$P'_n(k) = \begin{cases} 1, & \text{if } T_k a_{n,i} \in A_{n,j}, \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

Clearly, the computational effort involved in the numerical construction of $P'_n(k)$ is less than that of $P_n(k)$ in (31), especially in higher dimensions (in the Tips and Tricks section, we discuss other numerical methods of computing $P_n(k)$). We do not recommend using $P'_n(k)$ for deterministic systems, as the results are usually very poor. However, for random systems, one can still obtain quite good results with the cruder approximation of (32).

How these matrices are combined depends on whether the stochastic process is iid or Markov.

iid case In the iid case, we set

$$P_n = \sum_{k=1}^r w_k P_n(k) \tag{33}$$

Markov case In the Markov case, let W be the transition matrix for the Markov process that generates the random sequence of indices for the maps $\{T_k\}$.

Now set

$$S_n = \begin{pmatrix} W_{11}\ P_n(1) & W_{12}\ P_n(1) & \cdots & W_{1r}\ P_n(1) \\ W_{21}\ P_n(2) & W_{22}\ P_n(2) & \cdots & W_{2r}\ P_n(2) \\ \vdots & \vdots & \ddots & \vdots \\ W_{r1}\ P_n(r) & W_{r2}\ P_n(r) & \cdots & W_{rr}\ P_n(r) \end{pmatrix}, \tag{34}$$

In both the iid and Markov cases, the matrices P_n and S_n may be thought of as finite-dimensional projections of an "averaged" Perron-Frobenius operator; see [17] for details. With either of these two matrices, one may apply the methods described in §3 to estimate invariant objects such as invariant measures, invariant sets, Lyapunov exponents, and recurrence times.

4.2 Invariant measures

iid case We calculate the fixed left eigenvector p_n of P_n as constructed in (33), and normalise so that

$$\sum_{i=1}^{n} p_{n,i} = 1. \tag{35}$$

Set $\mu_n(A_{n,i}) = p_{n,i}$ and define the approximate invariant measure as in (11).

Markov case We calculate the fixed left eigenvector of S_n, and denote this as $s_n = [s_n^{(1)}|s_n^{(2)}|\cdots|s_n^{(r)}]$ where each $s_n^{(k)}$, $k = 1, \ldots, r$ is a vector of length n, and $\sum_{k=1}^{r}\sum_{i=1}^{n} s_{n,i}^{(k)} = 1$. Define the approximate invariant measure as

$$\mu_n(A_{n,i}) = \sum_{k=1}^{r} s_{n,i}^{(k)} \tag{36}$$

We now use (11) again to define a measure on all of M.

Results parallel to those of Theorem 3.1 hold for our random systems.

Proposition 4.6: *Suppose that each $T_k : M \circlearrowleft$ is continuous and the resulting random dynamical system has a physical measure μ (in the weaker sense where only (30) need hold, rather than (29)). Let $\{\mu_n\}$ denote a sequence of approximate invariant measures as defined in either (35) or (36) above, and let μ^* be a weak limit of this sequence. Denote by S the intersection $\bigcap_{n=n_0}^{\infty} \operatorname{supp} \mu_n$. Then the conclusions of Theorem 3.1 hold.*

PROOF: One first requires the facts that the matrices (33) and (34) represent a finite-dimensional approximation of an appropriately averaged Perron-Frobenius operator; this is detailed in [17][13]. With this established, the proofs run along the same lines as the deterministic case. $\square$

Example 4.7 (The (non-uniform) bouncing ball): We now suppose that our bouncing ball has gone soft on one side, so that sometimes we register a value of $\alpha = 0.1$, rather than the original value of $\alpha = 0.5$. We assume that every time it lands on the soft side, it will surely land on the good side next time, while if it lands on the good side, it has a 50/50 chance of landing on the soft side next time. The situation we describe is a Markov random composition of two mappings $T_{\alpha=0.5}$ and $T_{\alpha=0.1}$. The transition matrix for this Markov chain is $W = \begin{pmatrix} 1/2 & 1/2 \\ 1 & 0 \end{pmatrix}$, where $\alpha = 0.5$ is identified with state #1 and $\alpha = 0.1$ with state #2. We construct S_n as in (34), and compute the approximate invariant measure as in (36); see Figure 10. Again, there is good agreement between the two figures.

[13]The form of (34) is slightly different to the matrix given in [17] as we have performed a similarity transformation on the latter to yield a more intuitive representation.

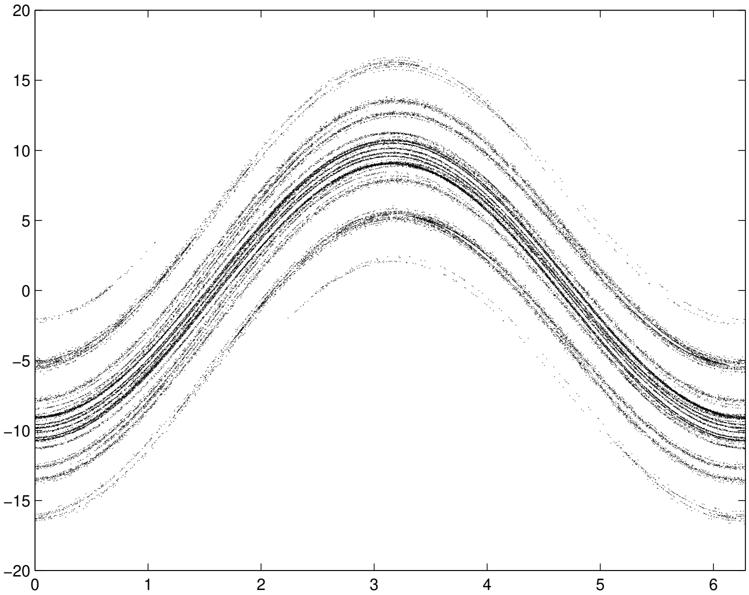

Figure 9: Plot of orbit of length 50000 for the random bouncing ball map

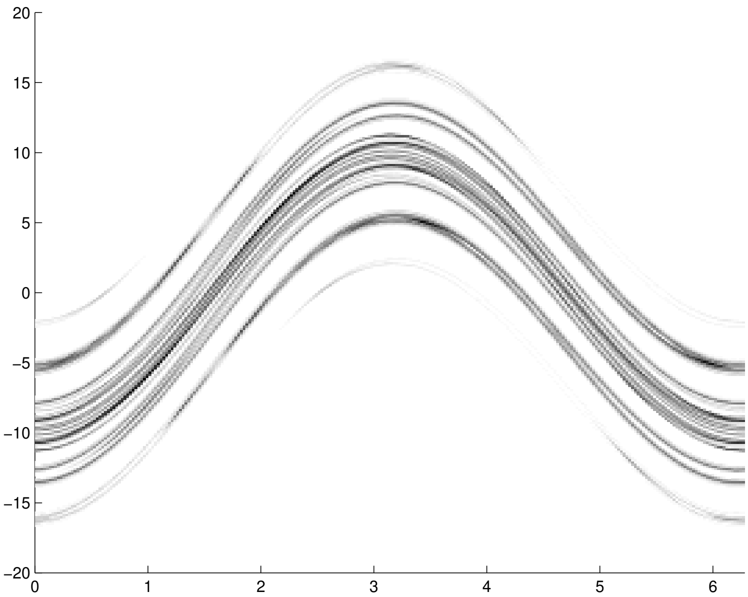

Figure 10: Approximate invariant measure for the random bouncing ball map using 18512 partition sets

4.2.1 Rigorous results

In certain situations, one is able to obtain rigorous upper bounds for the difference between μ_n and μ.

The first of these is where the random system *contracts* phase space on average. Typical examples of such systems are the iterated function systems (IFS's) of Barnsley [3] and co-workers, where a finite number of mappings are selected using an iid law to create fractal images. Suppose that T_k is selected with probability w_k, and define $s_k = \max_{x,y \in M} \|T_k x - T_k y\| / \|x - y\|$ as the Lipschitz constant for T_k; then set $s = \sum_{k=1}^{r} w_k s_k$. It is straightforward to show that if $s < 1$, then this random dynamical system has a unique invariant measure, the support of which is a fractal set. Furthermore, one has the bound[14] [20] (see also [55])

$$d_H(\mu, \mu_n) \leq \frac{1+s}{1-s} \max_{1\leq i\leq n} \operatorname{diam}(A_{n,i}), \tag{37}$$

where d_H is the natural metric generating the weak topology on measures, defined by $d_H(\nu_1, \nu_2) = \sup\left\{\left|\int_M h\ d\nu_1 - \int_M h\ d\nu_2\right|; h : M \to \mathbb{R} \text{ has Lipschitz constant } 1\right\}$. If a uniform grid is used, this bound may be improved by a factor of 2. Similar results hold for Markov compositions. Ulam-type methods of approximating invariant measures of IFS's are also discussed in the book [51].

The second situation is where the dynamical system is *expanding* on average. This setting is more complicated as the random system may have infinitely many invariant measures, and it is important to show that Ulam's method approximates the physical measure (in this expanding case, the physical measure will have a bounded density). In the case of iid random dynamical systems on the interval $[0, 1]$ where each map T_k is a *Lasota-Yorke map*, it is known that a bounded invariant density exists provided that $\sum_{k=1}^r w_k/|T_k'(x)| < 1$; see [50]. Under some additional conditions, it is shown in [17] that (i) the random system (either iid or Markov) possesses a unique bounded invariant density, and (ii) that the Ulam estimates μ_n converge to the physical measure μ (which has a bounded density). In addition, convergence rates of $O(\log n/n)$ for the difference $\|\mu - \mu_n\|_{L^1}$ are proven, and if each T_k is a C^2 map of the circle S^1, rather than of the interval $[0, 1]$, explicitly calculable numerical bounds for the error $\|\mu - \mu_n\|_{L^1}$ are given. In the future, we will no doubt see extensions of these results to higher dimensions.

4.3 Lyapunov exponents

The random version of (8) is

$$\lambda := \lim_{N\to\infty} \frac{1}{N} \log \|D_{x_{N-1}} T_{k_{N-1}} \circ \cdots \circ D_{x_1} T_{k_1} \circ D_x T_{k_0}(v)\|. \tag{38}$$

Often, the same value of λ is obtained for $\mathbb{P}$ almost all random sequences, Lebesgue almost all $x \in M$, and for *every* $v \in \mathbb{R}^d$. We denote this value by λ^1.

Things are very simple in the case of one-dimensional systems driven by an iid process. In this case, the expression (38) may be alternately expressed as

$$\lambda = \sum_{k=1}^r w_k \int_M \log |T_k'(x)|\ d\mu(x) \tag{39}$$

by a straightforward application of (29) with $f(\omega, x) = \log |T_{\omega_0}'(x)|$. Thus, once we have an estimate of the physical measure μ, it is easy to compute an approximation of λ via

$$\lambda_n := \sum_{k=1}^r w_k \int_M \log |T_k'(x)|\ d\mu_n(x) \approx \sum_{k=1}^r w_k \sum_{i=1}^n \log |T_k'(x_{n,i})| \cdot p_{n,i}, \tag{40}$$

where $x_{n,i} \in A_{n,i}$ (for example, the midpoint of the interval $A_{n,i}$).

[14] This rigorous result holds even when the crude approximation of Remark 4.5 is used.

4.3.1 Rigorous results

For random systems, we adopt Approach #2 of §3.3, as the other two approaches are not so helpful in the random situation. Here we only briefly describe the calculation of Lyapunov exponents of iid random dynamical systems where the Jacobian matrices are constant. This situation arises when each of the mappings T_k are affine (as with most IFS's). Equation (38) now becomes independent of the $x_0, \ldots, x_{N-1}$, and is a function only of the sequence $k_0, \ldots, k_{N-1}$; we are essentially dealing with an iid random matrix product.

Suppose that M is two-dimensional so that our Jacobian matrices are 2×2 matrices. We need to define a probability measure ξ on the angle space $\mathbb{RP}^1 \cong [0, \pi)$ (ξ is a relative of the probability measure alluded to in §3.3.2). Each matrix DT_k (note independence of x) defines a map from $[0, \pi)$ to itself via angle(v) = angle$(DT_k(v))$, where angle(v) is the angle between v and some fixed reference vector in $\mathbb{R}^2$. To simply notation, we will identify a vector $v \in \mathbb{R}^2$ and its angle with respect to some fixed reference vector. The Jacobian matrix DT_k will then be thought of as an action on the space of angles $[0, \pi)$.

Example 4.8: Suppose $DT_k = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$, and $v = (\,1 \quad 1\,)$. We identify v with the angle $\pi/2$ (this is the angle v makes with the reference vector $(\,1 \quad 0\,)$). Then $DT_k(v) = (\,1 \quad 3\,)$, and we identify this vector with the angle $\tan^{-1} 3$. By a slight abuse of notation, we may write $DT_k(\pi/2) = \tan^{-1}(3)$, and in this way we consider DT_k to be an action on the set of angles $[0, \pi)$.

The probability measure ξ on $\mathbb{RP}^1$ that we seek should satisfy:

$$\xi(E) = \sum_{k=1}^{r} w_k DT_k^{-1}(E) \tag{41}$$

for every measurable subset $E \subset [0, \pi)$. This is because of the following result.

Theorem 4.9 ([25]): *Suppose that each $d \times d$ matrix DT_k, $k = 1, \ldots, r$ is non-singular, and that the only subspace of $\mathbb{R}^d$ that is invariant under all of the DT_k is the trivial subspace. Then with probability one,*

$$\lambda^1 = \lim_{N \to \infty} \frac{1}{N} \log \|DT_{k_{N-1}} \circ \cdots \circ DT_{k_0}(v)\|$$

for every $v \in \mathbb{R}^d$. Furthermore,

$$\lambda^1 = \lambda(\xi) := \sum_{k=1}^{r} w_k \int_{\mathbb{RP}^{d-1}} \log \|DT_k(v)\| \, d\xi(v)$$

where ξ is any probability measure satisfying (41).

We approximate a measure ξ satisfying (41) in essentially the same way that we have already used for measures on M. Partition $[0, \pi)$ into a finite collection of intervals $E_1, \ldots, E_m \subset [0, \pi)$, and define an $m \times m$ stochastic matrix by:

$$D_{m,gh}(k) = \frac{m(E_g \cap DT_k^{-1}(E_h))}{m(E_g)} \tag{42}$$

Alternatively (*ala* Remark 4.5), one could choose a collection of points $e_1, \ldots, e_m$ such that $e_g \in E_g$ and set

$$D_{m,gh}(k) = \begin{cases} 1, & \text{if } DT_k(e_g) \in E_h \\ 0, & \text{otherwise.} \end{cases} \tag{43}$$

or use the other suggestions in §6.1. One now computes the fixed left eigenvector of the matrix $D_m = \sum_{k=1}^r w_k D_m(k)$; we denote this eigenvector by ξ_m. Selecting points $e_1, \ldots, e_m$ as before, we define an approximation of λ^1 as

$$\lambda_m^1 := \sum_{k=1}^r w_k \sum_{g=1}^m \log \|D_m(k)(e_g)\| \tag{44}$$

where, in the expression $\|D_m(k)(e_g)\|$, e_g is a unit vector in the direction represented by e_g, and we measure the length of the vector $D_m(k)(e_g)$ (this is the factor by which $D_m(k)$ stretches vectors in the direction of e_g). The above constructions may be generalised to higher dimensions; see [20] for details. We have summarised the simplest situation here; the treatment of Markov random matrix products, and iid random nonlinear dynamical systems may be found in [20].

4.4 Mean and variance of return times

To estimate the mean and variance of return times, we again construct a finite Markov model, and calculate the mean and variance of return times to a suitable set of states. In the iid case we can define a Markov model using (33), and proceed as for deterministic systems.

In the Markov case, we use (34), and produce a left eigenvector s_n of S_n such that $\sum_{k=1}^r \sum_{i=1}^n s_{n,i}^{(k)} = 1$. When writing S_n in the block form (25), recall that each partition set $A_{n,i}$ corresponds to r states of the Markov chain governed by S_n. With this in mind, one may substitute S_n and s_n into the algorithm described in §3.4. It is also possible to consider situations where the set B depends on the map T_k which is currently applied [19].

Example 4.10 (The non-uniform bouncing ball (cont...)): We return to the random dynamical system of Example 4.7 and compute the return times to low impact velocity configurations described by the set $S^1 \times [-1.25, 1.25]$.

Table 5: Estimates of the mean and variance of return times to the set $B = S^1 \times [-1.25, 1.25]$

Number of partition sets n	1738	5488	18512
Mean $\mathcal{M}_n$	11.18	10.93	10.98
Root Variance $\sqrt{\mathcal{V}_n}$	10.73	10.45	10.49

Compare these values with 11.05 ± 0.03 and 10.57 ± 0.03, the mean and variance respectively, obtained directly from 10 simulated random orbits of length 10^6 (plus/minus one standard deviation of the 10 values obtained). This example is one situation where we would benefit by using one of the alternate partition selection techniques described in §6.

4.5 Advantages for Markov modelling of random dynamical systems

To close this section, we discuss two further advantages of Markov modelling over trajectory simulation that are not present in the deterministic case.

The first of these concerns the accuracy of the approximations. For deterministic systems, the two competing approaches (simulating long orbits and coarse graining) both have their inaccuracies. The iterative approach of following long orbits (let's assume that we can do perfect computations) has the problem of temporal deviation from equilibrium behaviour. That is, we should really have orbits of *infinite* length, but instead we have orbits of *finite* length whose statistical behaviour is not the same. In contrast, with the Markov modelling approach, we can *exactly* compute the long term behaviour of our model, but we compute the long term behaviour of an *approximation* of the system, rather than that of the true system. Turning now to random systems, the iterative approach is fighting errors on two fronts; namely, the deviation from equilibrium in phase space mentioned above, and additionally, the deviation of the distribution of the finite length random sequence of indices $k_0, \ldots, k_{N-1}$ from its equilibrium distribution $\mathbb{P}$. On the other hand, the Markov modelling approach completely eliminates errors from the random fluctuations by averaging them out through the expectations performed in its construction. Thus our Markov models for random dynamical systems do not suffer from the inaccuracies due to random fluctuations, and are therefore (heuristically at least) more accurate; this is borne out numerically in Lyapunov exponent computations [20].

The second advantage lies in the flexibility of the Markov modelling approach regarding the underlying stochastic process. Suppose that we wish to study a *family of systems* which use the same maps $T_1, \ldots, T_r$, but a different distribution $\mathbb{P}$ (in the bouncing ball example, this would amount to varying the probabilities with which impacts occur on the soft and hard sides). Most of the computational effort goes into constructing the (fixed) transition matrices $P_n(k)$, $k = 1, \ldots, r$, while the ancillary calculations involving eigenvectors and so on, are relatively cheap. Thus, we may perform analyses on a *whole family of systems* very quickly, by reusing most of the original constructions. In contrast, if we were to use the direct method of simulating long orbits, then entirely new orbit calculations would be required for each new set of probabilities.

5 Miscellany

We briefly outline some other applications and techniques related to our approach of Markov modelling. Unless otherwise stated, we refer only to deterministic systems.

Global attractors Related partition-based methods may be used to approximate the global attractor for a given subset of phase space. If $B \subset M$, then the global attractor of B is defined by $G = \bigcap_{k\geq 0} T^j(B)$. Methods of computing an (in principle[15]) rigorous box covering of the global attractor are detailed in [8]. Bounds for the Hausdorff distance between the approximate covering and the global attractor are given for uniformly hyperbolic diffeomorphisms.

Using similar techniques, box coverings for global attractors $G(\omega)$ of individual sample paths ω of random dynamical systems have been studied [38].

[15] when combined with Lipschitz estimates for the map [33]

Work is in progress on covering "averaged" global attractors of random systems; such global attractors contain all possible random orbits and so $G_{av} = \bigcup_\omega G(\omega)$.

Noisy systems A popular idea is to "noise up" a system by considering the transformation $x \mapsto Tx + \epsilon$, where the perturbation $\epsilon \in \mathbb{R}^d$ is chosen in a uniform distribution from some small ball centred around 0. This may be viewed as defining a random dynamical system where the collection of maps $T_\epsilon = Tx + \epsilon$ are applied in an iid fashion with equal probability. The Perron-Frobenius operator $\mathcal{P}_\epsilon : L^2(M, m) \circlearrowleft$ for this random perturbation has the desirable property that it is a compact operator under very mild conditions on T, and this greatly simplifies the ergodic theoretic analysis. For example, it is relatively easy to show that this noisy system has a unique invariant probability density, in contrast to the purely deterministic case. The finite-state Markov modelling may now be applied to the perturbed system, and various convergence results proven concerning the approximation of invariant measures and invariant sets; see [11]. This setting forms the basis of the thesis [32], where the merits of alternative partitioning methods are also considered.

Rotation numbers The approximation of rotation numbers of orientation preserving C^2 circle diffeomorphisms using Ulam constructions is described in [57].

Topological entropy It is possible to obtain rigorous upper bounds for the topological entropy of T with respect to a fixed (coarse) partition. All orbits of T are possible under the Markov model, however the converse is not true. In this sense, the Markov model is more "complex" from the orbit generating viewpoint. However, as the partitions are refined and our Markov model becomes more accurate, these extra orbits are successively eliminated, so that our upper bounds become increasingly sharp. This is work in progress [24].

Spectra of "averaged" transfer operators for random systems One may also attempt to garner dynamical information from the spectrum and eigenvectors of the matrices (33) and (34), in analogy to the deterministic case. This is work in progress.

6 Numerical Tips and Tricks

We discuss methods of computing the transition matrix and of partition selection. Most transition matrix computations in this chapter have used the GAIO software package, available on request from `http://www-math.uni-paderborn.de/~agdellnitz/gaio/`. Algorithms (ii) and (iii) of §6.1 and (i)–(iv) of §6.2 have been coded in this software.

6.1 Transition matrix construction

Techniques for the computation of the transition matrix may be split into three main classes; namely "exact" methods, Monte-Carlo/Imaging methods, and an exhaustive method of approximation.

(i) *"Exact" methods*: For one-dimensional systems, it is often possible to construct the transition matrix exactly. If the map is locally one-to-one on each partition set, then only the inverse images of the endpoints of each set need be calculated.

If the inverse images are difficult to obtain, an alternative is to compute the matrix

$$P'_{n,ij} = \frac{m(TA_i \cap A_j)}{m(TA_i)}, \tag{45}$$

which in one-dimension again requires only the computation of forward images of partition endpoints.

The matrix P'_n is not useful theoretically because (i) forward images of sets may not be measurable, while inverse images (T continuous) of Borel measurable sets are always measurable, and (ii) the standard form (3) arises as a discretisation of the Perron-Frobenius operator for T, while (45) does not. If T is linear on each A_i, then $P_{ij} = P'_{ij}$ for $j = 1, \ldots, n$; this forward-imaging exact computation was carried out for two-dimensional piecewise linear reconstructions in [23]. Otherwise, the difference between P and P' is governed by the second derivative of T and the diameter of the partition sets. We do not recommend using forward-imaging for maps with very large second derivatives.

(ii) *Monte-Carlo / Imaging of test points*: The most popular method is the so-called Monte-Carlo approach [31]. To compute P_{ij}, one randomly selects a large number of points $\{a_1, \ldots, a_N\} \subset A_i$, and sets $P_{ij} \approx \#\{a \in \{a_1, \ldots, a_N\} : T(a) \in A_j\}/N$. A similar approach is to choose a uniform grid of test points within each partition set, and perform the same calculation. My personal feeling is that the latter approach is better as the uniform grid more reliably approximates Lebesgue measure. Any number of variations on the selection of test points can be used, though Monte-Carlo and uniform grids are the most common.

(iii) *Exhaustion*: A recent approach [29] is to rigorously approximate the transition probability by a means of *exhaustion* akin to the exhaustion methods of Eudoxus. To compute the Lebesgue measure of the portion of A_i that is mapped into A_j, one repeatedly refines the set A_i until it is known (via Lipschitz estimates on the map) that a refined subset of A_i is mapped *entirely inside* A_j. In this way, the set A_i is repeated broken down into small pieces which map entirely inside A_j, with this process terminating at the desired level of precision.

6.2 Partition selection

This section is devoted to suggesting methods of producing better Markov models via smarter partition selection. That is, how should one choose partitions to best capture the dynamics of the system.

Of course, if a Markov partition is available, this is clearly the best choice. However, we are assuming that this is not the case, and we are left with the decision of how to construct a suitable "grid". For the most part, we consider partition selection where the criteria for a good partition is that it produces a good estimate of the physical measure (at least a better estimate than a uniform grid would produce). Of course, often we don't know what the physical measure is, and so this mostly restricts rigorous numerical testing to one-dimensional systems. Nevertheless, we outline three main approaches, and suggest heuristically when they may be useful. In all cases, one selects an initial coarse partition, computes the invariant measure for the Markov model, and on the basis of information contained in the invariant measure of the current model, a choice is made on which partition sets to refine and which to not refine.

(i) *Standard approach*: Refine any partition sets which are currently assigned non-zero measure.

(ii) *Equal mass approach*: Refine any partition sets which are assigned a measure greater than $1/n$, where n is the current number of partition sets; [10].

The rationale behind this is that one should focus more finely on regions where there is large mass. In this sense, the method is not only targeted at obtaining more accurate estimates of the invariant measure, but also more accurate modelling of the *dynamics* of the system; this has been demonstrated for the estimation of return times in [19]. This method is particularly suited to systems which possess a singular physical measure, as in dissipative chaotic systems. However, because of the non-uniform refinement (the minimal ratio of cell sizes is almost always at least 2 if refinement is done by "halving" a set), it often performs worse than the standard method in cases where the physical measure is smooth. For this approach to be useful, the ratio $\sup_{x\in\mathrm{supp}\,\mu_n}\phi_n(x)/\inf_{x\in\mathrm{supp}\,\mu_n}\phi_n(x)$ should be much larger than 2 for all $n\geq 0$ (ϕ_n is the density of the approximate measure μ_n).

(iii) *High derivative approach*: Let c_i denote the "centre point" of a partition set A_i. Refine partition sets where the value of

$$E_i := m(A_i)\,\mathrm{diam}(A_i)\max\left\{\frac{|p_i/m(A_i)-p_j/m(A_j)|}{|c_i-c_j|} : A_j \text{ is a neighbour of } A_i\right\}, \quad (46)$$

is greater than $(1/n)\sum_{i=1}^n E_i$; [30].

The expression that is maximised over is meant to be an approximation of the derivative of the invariant density ϕ on A_i in the "direction of" A_j. In [30], one assumes that the physical measure is smooth; therefore, if the current estimate of the invariant measure has adjacent sets given very different measures, there must be an error in this region, and so one refines these sets to obtain better estimates. The number E_i is intended to approximate the error incurred on the partition set A_i.

An alternative viewpoint is as follows. In [16], it is noted that the matrix $\tilde{P}_{ij} := \mu(A_i\cap T^{-1}A_j)/\mu(A_i)$ is an optimal approximation in the sense that the fixed left eigenvector $\tilde{p}$ of $\tilde{P}$ assigns exactly the correct weights to the partition sets; that is, $\tilde{p}_i=\mu(A_i)$ (this approach is also followed in [35]). The difference between P and $\tilde{P}$ is essentially given by how "non-Lebesgue-like" the measure μ is *within each partition set*; roughly speaking, how "non constant" the distribution of μ is within partition sets. One may try to reduce[16] the error $\|\mu-\mu_n\|_1$ by creating a partition which produces a transition matrix P similar to that of the special matrix $\tilde{P}$. Such an analysis also leads to the error minimisation criteria (46).

The high derivative method is targeted specifically towards more accurate estimates of the physical measure. It often performs better than the equal mass approach for maps with smooth densities.

(iv) *Large difference approach* [32]: One refines all partition sets and constructs a temporary transition matrix P_{temp} and invariant measure p_{temp} for the refined partition. This refined invariant measure is compared with the invariant measure p_{old} and only sets in the old partition for which the measure according to p_{temp} and p_{old} is very different are

[16]Such a "distortion reducing" approach is also discussed in [47] for one-dimensional maps (in particular, the logistic family $T_ax=ax(1-x)$), and a relative of the *Equal mass approach* is advocated as a means to make P better approximate $\tilde{P}$.

split up. The transition matrix P_{temp} is now discarded. This approach is based on a standard method of numerical analysis.

Comparisons of the three alternative approaches are detailed in [32].

7 Acknowledgements

We thank Oliver Junge for helpful discussions and critical comments during the preparation of this chapter, and Oliver Junge and Stefan Sertl for assistance with the use of the GAIO package. The list of references is not meant meant to be exhaustive, and is far from being so. GF apologises for any oversights.

References

[1] Ludwig Arnold. *Random Dynamical Systems.* Springer monographs in mathematics. Springer, Berlin, 1998.

[2] Philip J. Aston and Michael Dellnitz. The computation of Lyapunov exponents via spatial integration with application to blowout bifurcations. *Computer methods in applied mechanics and engineering*, 170:223–237, 1999.

[3] Michael F. Barnsley. *Fractals everywhere.* Academic Press, Boston, 1988.

[4] George D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences of the USA*, 17:656–60, 1931.

[5] J. R. Blum and J. I. Rosenblatt. On the moments of recurrence time. *Journal of Mathematical Sciences*, 2:1–6, 1967.

[6] Abraham Boyarsky and Pawel Góra. *Laws of Chaos – Invariant Measures and Dynamical systems in One Dimension.* Probability and Its Applications. Birkhäuser, Boston, 1997.

[7] Michael Dellnitz, Gary Froyland, and Stefan Sertl. On the isolated spectrum of the Perron-Frobenius operator. Submitted to *Nonlinearity*, 1999.

[8] Michael Dellnitz and Andreas Hohmann. A subdivision algorithm for the computation of unstable manifolds and global attractors. *Numerische Mathematik*, 75:293–317, 1997.

[9] Michael Dellnitz and Oliver Junge. Almost invariant sets in Chua's circuit. *International Journal of Bifurcation and Chaos*, 7(11):2475–2485, 1997.

[10] Michael Dellnitz and Oliver Junge. An adaptive subdivision technique for the approximation of attractors and invariant measures. *Computing and Visualization in Science*, 1:63–68, 1998.

[11] Michael Dellnitz and Oliver Junge. On the approximation of complicated dynamical behavior. *SIAM Journal on Numerical Analysis*, 36(2):491–515, 1999.

[12] Jiu Ding and Aihui Zhou. Finite approximations of Frobenius-Perron operators. a solution of Ulam's conjecture to multi-dimensional transformations. *Physica D*, 92:61–68, 1996.

[13] H. Francke, D. Plachky, and W. Thomsen. A finitely additive version of Poincaré's recurrence theorem. In N. Christopeit, K. Helmes, and M. Kohlmann, editors, *Stochastic Differential Systems - Proceedings of the 3rd Bad Honnef Conference, June 3–7, 1985*, Lecture Notes in Control and Information Science, Berlin, 1985. Springer-Verlag.

[14] Gary Froyland. Finite approximation of Sinai-Bowen-Ruelle measures of Anosov systems in two dimensions. *Random & Computational Dynamics*, 3(4):251–264, 1995.

[15] Gary Froyland. Computer-assisted bounds for the rate of decay of correlations. *Communications in Mathematical Physics*, 189(1):237–257, 1997.

[16] Gary Froyland. Approximating physical invariant measures of mixing dynamical systems in higher dimensions. *Nonlinear Analysis, Theory, Methods, & Applications*, 32(7):831–860, 1998.

[17] Gary Froyland. Ulam's method for random interval maps. *Nonlinearity*, 12(4):1029–1052, 1999.

[18] Gary Froyland. Using Ulam's method to calculate entropy and other dynamical invariants. *Nonlinearity*, 12:79–101, 1999.

[19] Gary Froyland and Kazuyuki Aihara. Estimating the mean and variance of interspike intervals. In preparation.

[20] Gary Froyland and Kazuyuki Aihara. Rigorous numerical estimation of Lyapunov exponents and invariant measures of iterated function systems and random matrix products. *Int. J. Bifur. Chaos.* 10, 2000. To appear.

[21] Gary Froyland, Kevin Judd, and Alistair I. Mees. Estimation of Lyapunov exponents of dynamical systems using a spatial average. *Physical Review E*, 51(4):2844–2855, 1995.

[22] Gary Froyland, Kevin Judd, Alistair I. Mees, and Kenji Murao. Lyapunov exponents and triangulations. In *Proceedings of the 1993 International Symposium on Nonlinear Theory and its Applications, Hawaii, December 1993*, volume 1, pages 281–286, 1993.

[23] Gary Froyland, Kevin Judd, Alistair I. Mees, Kenji Murao, and David Watson. Constructing invariant measures from data. *International Journal of Bifurcation and Chaos*, 5(4):1181–1192, 1995.

[24] Gary Froyland, Oliver Junge, and Gunter Ochs. Rigorous computation of topological entropy of multidimensional dynamical systems with respect to a finite partition. In preparation.

[25] Harry Furstenberg and Yuri Kifer. Random matrix products and measures on projective spaces. *Israel Journal of Mathematics*, 46(1–2):12–32, 1983.

[26] Pawel Góra. On small stochastic perturbations of mappings of the unit interval. *Colloquium Mathematicum*, 49(1):73–85, 1984.

[27] Pawel Gora and Abraham Boyarsky. Absolutely continuous invariant measures for piecewise expanding C^2 transformations in R^N. *Israel Journal of Mathematics*, 67(3):272–286, 1989.

[28] John Guckenheimer and Philip Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, volume 42 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983.

[29] R. Guder, M. Dellnitz, and E. Kreuzer. An adaptive method for the approximation of the generalized cell mapping. *Chaos, Solitons, and Fractals*, 8(4):525–534, 1997.

[30] Rabbijah Guder and Edwin Kreuzer. Adaptive refinement of the generalized cell mapping. Preprint. 1998.

[31] Fern Y. Hunt. A Monte Carlo approach to the approximation of invariant measures. *Random & Computational Dynamics*, 2(1):111–133, 1994.

[32] Oliver Junge. *Mengenorientierte Methoden zur numerischen Analyse dynamischer Systeme*. PhD thesis, University of Paderborn, Paderborn, 1999.

[33] Oliver Junge. Rigorous discretization of subdivision techniques. In *Proceedings of Equadiff '99, Berlin, August 1999.*, 1999.

[34] M. Kac. On the notion of recurrence in discrete stochastic processes. *Bulletin of the American Mathematical Society*, 53:1002–1010, 1947.

[35] Michael Keane, Rua Murray, and Lai-Sang Young. Computing invariant measures for expanding circle maps. *Nonlinearity*, 11(1):27–46, 1998.

[36] Gerhard Keller. Stochastic stability in some chaotic dynamical systems. *Monatshefte für Mathematik*, 94:313–353, 1982.

[37] Gerhard Keller and Carlangelo Liverani. Stability of the spectrum for transfer operators. Preprint, 1998.

[38] Hannes Keller and Gunter Ochs. Numerical approximation of random attractors. Institut für Dynamische Systeme, Universität Bremen, Report Nr. 431, August 1998.

[39] R.Z. Khas'minskii. Principle of averaging for parabolic and elliptic differential equations and for Markov processes with small diffusion. *Theory of Probability and its Applications*, 8(1):1–21, 1963.

[40] Yuri Kifer. *Ergodic Theory of Random Transformations*, volume 10 of *Progress in Probability and Statistics*. Birkhäuser, Boston, 1986.

[41] Yuri Kifer. *Random Perturbations of Dynamical Systems*, volume 16 of *Progress in Probability and Statistics*. Birkhäuser, Boston, 1988.

[42] Andrzej Lasota and Michael C. Mackey. *Chaos, Fractals, and Noise. Stochastic Aspects of Dynamics*, volume 97 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1994.

[43] Andrzej Lasota and James A. Yorke. On the existence of invariant measures for piecewise monotonic transformations. *Transactions of the American Mathematical Society*, 186:481–488, 1973.

[44] Tien-Yien Li. Finite approximation for the Frobenius-Perron operator. A solution to Ulam's conjecture. *Journal of Approximation Theory*, 17:177–186, 1976.

[45] Walter M. Miller. Stability and approximation of invariant measures for a class of nonexpanding transformations. *Nonlinear Analysis*, 23(8):1013–1025, 1994.

[46] Walter M. Miller and Fern Y. Hunt. Approximation of attractors for finite dimensional maps by Ulam's method. Preprint, May 1996.

[47] Rua Murray. Adaptive approximation of invariant measures. Preprint. 1998.

[48] Rua Murray. Existence, mixing and approximation of invariant densities for expanding maps on R^d. Preprint. 1998.

[49] Rua Murray. Approximation error for invariant density calculations. *Discrete and Continuous Dynamical Systems*, 4(3):535–557, 1998.

[50] S. Pelikan. Invariant densities for random maps of the interval. *Transactions of the American Mathematical Society*, 281(2):813–825, 1984.

[51] Mario Peruggia. *Discrete Iterated Function Systems*. A K Peters, Wellesley, 1993.

[52] Karl Petersen. *Ergodic Theory*, volume 2 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1983.

[53] David Ruelle. Ergodic theory of differentiable dynamical systems. *IHES Publicationes Mathematiques*, 50:275–320, 1979.

[54] Leonard A. Smith. The maintenance of uncertainty. In G. Cini Castagnoli and A. Provenzale, editors, *Past and Present Variability of the Solar-Terrestrial system: Measurement, Data Analysis and Theoretical Models*, volume CXXXIII of *Proceedings of the International School of Physics*, pages 177–246, Amsterdam, 1997. Italian Physical Society.

[55] Jaroslav Stark. Iterated function systems as neural networks. *Neural Networks*, 4:679–690, 1991.

[56] S. M. Ulam. *Problems in Modern Mathematics*. Interscience, 1964.

[57] Bodo Werner and N. Nicolaisen. Discretization of circle maps. *Zeitscrift für Angewandte Mathematik und Physik*, 49(6):869–895, 1998.

[58] J. Wolfowitz. The moments of recurrence time. *Proceedings of the American Mathematical Society*, 18:613–614, 1967.

[59] Lai-Sang Young. Recurrence times and rates of mixing. To appear in *Israel Journal of Mathematics*.