



RIGOROUS NUMERICAL ESTIMATION OF LYAPUNOV EXPONENTS AND INVARIANT MEASURES OF ITERATED FUNCTION SYSTEMS AND RANDOM MATRIX PRODUCTS

GARY FROYLAND*

*Department of Mathematics and Computer Science, University of Paderborn,
Warburger Str. 100, Paderborn 33098, Germany*

KAZUYUKI AIHARA

*Department of Mathematical Engineering and Information Physics,
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

Received January 5, 1999; Revised April 13, 1999

We present a fast, simple matrix method of computing the unique invariant measure and associated Lyapunov exponents of a nonlinear iterated function system. Analytic bounds for the error in our approximate invariant measure (in terms of the Hutchinson metric) are provided, while convergence of the Lyapunov exponent estimates to the true value is assured. As a special case, we are able to rigorously estimate the Lyapunov exponents of an iid random matrix product. Computation of the Lyapunov exponents is carried out by evaluating an integral with respect to the unique invariant measure, rather than following a random orbit. For low-dimensional systems, our method is considerably quicker and more accurate than conventional methods of exponent computation. An application to Markov random matrix product is also described.

1. Outline and Motivation

This paper is divided into three parts. Firstly, we consider approximating the invariant measure of a contractive iterated function system (the density of dots that one sees in computer plots). Secondly, we describe a method of estimating the Lyapunov exponents of an affine IFS, or equivalently, the Lyapunov exponents of an iid random matrix product. While invariant measures of iterated function systems and Lyapunov exponents of random matrix products may seem rather like different objects, they are mathematically very similar, and our method of approximation is almost identical for each. In both cases, we discretise appropriate spaces, as an alternative to random itera-

tion. Finally, we bring the invariant measure and Lyapunov exponent approximations together to produce a method of estimating the Lyapunov exponents of a nonlinear IFS.

1.1. *Invariant measures of iterated function systems*

Contractive iterated function systems are relatively simple from the point of view of ergodic theory, as they have a unique invariant measure. We call a (Borel) probability measure μ *invariant* if $\mu = \sum_{k=1}^r w_k \mu \circ T_k^{-1}$, where the weights w_k and maps T_k , $k = 1, \dots, r$ define our IFS (formal definitions appear later). This invariant measure is the

*E-mail: froyland@math.uni-paderborn.de; URL: <http://www-math.uni-paderborn.de/~froyland>

distribution that you would “see” if you iterated an initial “blob” of mass forward for an infinitely long time. It is also the distribution that appears when random orbits of iterated function systems are plotted on a computer. As there is only one invariant measure, it describes the distribution of *almost all* [Elton, 1987] random trajectories of the IFS, and so plays a commanding role in the behavior of the dynamics. However, despite its ubiquity, methods for obtaining its numerical approximation have been rather rudimentary to date.

The most common way of obtaining a numerical approximation is to form a histogram from a very long random orbit. The IFS is defined by a finite collection of maps T_1, \dots, T_r . At each time step, a map T_k is chosen in an iid fashion and applied to the current point. In this way, long random orbits are produced. There is a theorem [Elton, 1987] that says that this method works, but it is often slow and there are no bounds on the error for finite length orbits. The authors know of two other constructions in the literature. Firstly, Boyarsky and Lou [1992] introduced a matrix method that is applicable only when each map T_k is Jablonski (in two dimensions, this means that each T_k has the form $T_k(x_1, x_2) = (T_{k,1}(x_1), T_{k,2}(x_2))$, where $T_{k,1}$ and $T_{k,2}$ are one-dimensional maps). Their result is very restrictive in the class of mappings it may be applied to, and their results do not include any error bounds for the approximation. Secondly, the book by Peruggia [1993] introduces a general method of discretisation, as a way of approximating the attracting invariant sets and the invariant measures of an IFS. While the constructions in [Peruggia, 1993] are similar to those of the present paper, our approach is entirely different, as the results of [Peruggia, 1993] rely on random iteration, which we wish to avoid. Because of this reliance on random iteration, [Peruggia, 1993] contains no quantitative bounds on the accuracy of the approximation.

In this paper, we present a simple computational method of rigorously approximating the unique invariant measure of an IFS (linear or nonlinear) based on a discretisation of the dynamics. As we are producing a computer estimate of the invariant measure, it must be in the form of a histogram, on a partition of the user’s choice. The number of calculations required is $O(n)$, where n is

the cardinality of the histogram partition, while the accuracy¹ is $O(n^{-1/d})$, where the invariant attracting set lies in \mathbb{R}^d . We present analytic bounds for the accuracy of our approximation in terms of the size of the histogram partition sets.

Remark 1.1. After completion of this work, the paper by Stark [1991] was brought to our attention, in which results similar to those of Theorem 2.2 were obtained with a view to implementation on a neural network.

1.2. *Lyapunov exponents of random matrix products and nonlinear iterated function systems*

We also describe a new rigorous numerical method of computing the Lyapunov exponents of an IFS. In analogy to the deterministic case [Benettin *et al.*, 1980], the traditional method of exponent computation is to choose a random starting point in phase space and run out a long random orbit

$$x_N = x_N(k_{N-1}, \dots, k_0, x_0) = T_{k_{N-1}} \circ \dots \circ T_{k_0} x_0, \quad (1)$$

where k_i , $i = 1, \dots, N - 1$ are iid random variables that determine which map from a finite collection is to be applied. The local rate of contraction in a randomly chosen direction is averaged along this orbit by sequentially applying the Jacobian matrices of the maps. Symbolically, the Lyapunov exponents λ are computed as a *time average* by:

$$\lambda := \lim_{N \rightarrow \infty} \log(\|DT_{k_{N-1}}(x_{N-1}) \circ \dots \circ DT_{k_0}(x_0)v\|^{1/N}), \quad (2)$$

where $DT_{k_i}(x_i)$ denotes the Jacobian matrix of the map applied at the i th iteration, evaluated at x_i (the i th element of the random time series), and v denotes a starting vector. It is numerically observed, and in some cases may be proven, that the same value of λ is obtained for *all* starting vectors v .

Even in simple situations, such as an IFS comprised of *affine* mappings, the standard time averaging technique may suffer from instabilities and random fluctuations. This is so especially if some of the Jacobian matrices are near to singular, or some of the mappings of the IFS are chosen with very small probability. Using a simple affine IFS, we

¹Using adaptive methods of subdividing the state space, the accuracy may be improved if a “thin” fractal set is living in a higher-dimensional Euclidean space; see [Dellnitz & Junge, 1998] for example.

demonstrate that Lyapunov exponent calculations via a time average, fluctuate significantly along the orbit, and even with long orbits the instabilities persist. Perhaps more unsettling is the observation that for very long orbits, the variation of exponent estimates *between individual orbits* is much greater than the variation along a particular orbit. Thus, apparent convergence to some value along a particular orbit need not imply that the value is a good estimate of a Lyapunov exponent.

Rather than be subjected to the random nature of the time average, we instead perform a *space average*, which makes direct use of our approximation of the invariant measure. The Lyapunov exponents are then calculated as an expectation or integral that automatically averages out the random fluctuations and is unaffected by near singular Jacobian matrices or infrequently applied maps. We show how to calculate all of the Lyapunov exponents of an IFS, be it linear or nonlinear.

We begin with the case where each mapping is affine, as we are here simply dealing with an iid random matrix product, where at each step a matrix is selected from a finite collection of matrices. We give a detailed example for a well-known affine IFS and compare our results with the standard time average. Later we demonstrate an extension of our method to a random matrix product where the matrices are selected according to a Markov process. In the case of nonaffine IFS's, our method is a generalization of the space averaging method for deterministic systems [Froyland *et al.*, 1995].

While we acknowledge that our methods do not provide analytic solutions for the Lyapunov exponents, they are mathematically rigorous, and in low-dimensional systems we have found them to provide estimates that are much more stable and accurate than those obtained via random iteration.

2. Invariant Measures of Iterated Function Systems

2.1. Background

The purpose of this paper is to numerically demonstrate the application of our methods and to compare and contrast them with standard techniques. Proofs of most results and proof sketches of the

remainder are given in the appendix. The theory for the more technical results is developed in [Froyland, 1998]. We begin by formalizing the mathematical objects that we will work with, so that we are able to state precise results. In the next section we give detailed examples of the use of our method.

Typically our IFS will act on some compact subset M of \mathbb{R}^d . One has a finite collection of Lipschitz mappings $T_1 : M \rightarrow M, \dots, T_r : M \rightarrow M$, and an associated collection of probabilities w_1, \dots, w_r such that $\sum_{k=1}^r w_k = 1$ and $w_k \geq 0, k = 1, \dots, r$. At each iteration step, a map T_k is selected with probability w_k , and applied to the current point in our space M . In this way, random orbits $\{x_i\}_{i=0}^{N-1}$ are generated as in (1).

We wish to describe the asymptotic distribution of such random orbits. Often, one obtains the same asymptotic distribution for every initial point $x \in M$ and almost every random orbit. A sufficient² condition for our IFS to possess only one such asymptotic distribution is that the IFS is a *contraction on average*; that is,

$$s := \sum_{k=1}^r w_k s_k < 1, \quad (3)$$

where

$$s_k := \sup_{x, y \in M} \frac{\|T_k x - T_k y\|}{\|x - y\|} \quad (4)$$

is the Lipschitz constant for T_k .

At this point, we make precise what is meant by an “asymptotic distribution”. Denote by $\mathcal{M}(M)$ the space of Borel probability measures on M ; elements of this space describe “mass” distributions on M . We seek a probability measure $\mu \in \mathcal{M}(M)$ satisfying

$$\mu = \sum_{k=1}^r w_k \mu \circ T_k^{-1}, \quad (5)$$

and will call such a probability measure μ an *invariant probability measure*. Later we shall use the operator $\mathcal{P} : \mathcal{M}(M) \rightarrow \mathcal{M}(M)$ defined by

$$\mathcal{P}\nu = \sum_{k=1}^r w_k \nu \circ T_k^{-1}, \quad \nu \in \mathcal{M}(M). \quad (6)$$

The invariance of a measure μ is equivalent to μ being a fixed point of \mathcal{P} .

²There are weaker conditions ([Barnsley & Elton, 1988] for example) that imply the existence of a unique invariant probability measure, but as we desire a *bound* for our approximation, we use the simple condition (3).

Remark 2.1. There are several reasons why we choose to define a measure μ satisfying (5) as “invariant” under our IFS. Firstly, recall that the measure $\mu \circ T_k^{-1}$ describes the image of the mass distribution μ under T_k . Thus, the RHS of (5) is a weighted average of the images of μ under the various mappings T_k . As such, μ has the interpretation of being invariant *on average* under the action of the IFS. Secondly, a theorem by Elton [1987] states that the measure μ is the unique probability measure that is “exhibited” by almost all orbits of the IFS. That is, if one defines a probability measure by

$$\tilde{\mu} = \tilde{\mu}(x, k_0, k_1, \dots) := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^{N-1} \delta_{x_i}, \quad (7)$$

where x_i is as in (1), and a weak limit (see below) is meant, then for every $x \in M$ and almost all sequences of map choices (k_0, k_1, \dots) , one has $\tilde{\mu} = \mu$. Thus the distribution of points in M that one sees from a finite random orbit of the IFS will almost always be close to the unique measure μ satisfying (5). Thirdly, the measure μ appears as a projection of an invariant measure of a deterministic representation of the IFS, called a *skew product*. This deterministic representation is used to formalize the mathematics from an ergodic theoretic point of view; see [Froyland, 1998] for a discussion.

In the sequel, we will be trying to approximate various measures, so we need a metric on $\mathcal{M}(M)$ in order to tell how good our approximations are, or if convergence occurs at all. The measures we will be attempting to approximate live on fractal sets and have a description which is infinite in nature, while our numerical approximations are in the form of histograms and have a finite description. For this reason, we cannot expect our computer approximations to be accurate in the sense of strong convergence; the natural choice for fractal measures is approximation in the sense of weak convergence (see [Lasota & Mackey, 1994, Sec. 12.2] for further details on strong and weak convergence of measures). The effect (and sometimes definition) of weak convergence is that:

$$\mu_n \rightarrow \mu \text{ weakly} \iff \int_M g d\mu_n \rightarrow \int_M g d\mu \quad \text{for all } g \in C(M, \mathbb{R}). \quad (8)$$

Thus if we are concerned with how our measures integrate continuous functions, then weak convergence is the natural choice for convergence. A very

useful property of weak convergence is that given a sequence of probability measures $\{\mu_n\}$ we may always find a weakly convergent subsequence. That is, the space $\mathcal{M}(M)$ is compact with the topology induced by weak convergence.

In order to quantify how good our approximations are, we require a metric. A useful metric which generates weak convergence as defined in (8) is the “Hutchinson metric”, defined by

$$d_H(\mu, \nu) = \sup_{h \in \text{Lip}(1)} \left| \int_M h d\mu - \int_M h d\nu \right|, \quad (9)$$

where $\text{Lip}(s)$ is the space of Lipschitz functions with Lipschitz constant s ; that is $\text{Lip}(s) = \{h \in C(M, \mathbb{R}) : |h(x) - h(y)| \leq s\|x - y\|\}$. For further details on weak convergence, see [Falconer, 1997].

2.2. Discretization and statement of result

To find a fixed point of the infinite-dimensional operator \mathcal{P} is difficult, so we replace \mathcal{P} with a simpler finite-dimensional operator \mathcal{P}_n whose fixed points are close to that of \mathcal{P} .

Construct a partition of M into n connected sets $\{A_1, \dots, A_n\}$. From each set, choose a single point $a_i, i = 1, \dots, n$ and for each mapping T_k define an $n \times n$ stochastic matrix $P_n(k)$ by setting

$$P_{n,ij}(k) = \begin{cases} 1, & \text{if } T_k a_i \in A_j, \\ 0, & \text{otherwise} \end{cases}. \quad (10)$$

This matrix requires $O(n)$ iterations to construct and is very sparse. Combine these r matrices to form the $n \times n$ matrix P_n :

$$P_n := \sum_{k=1}^r w_k P_n(k) \quad (11)$$

and denote by p_n a normalized fixed left eigenvector of P_n . We are now in a position to define our approximation of μ . We simply place a weight of $p_{n,i}$ at the position a_i , forming a probability measure that is a convex combination of n δ -measures. That is,

$$\mu_n := \sum_{i=1}^n p_{n,i} \delta_{a_i}. \quad (12)$$

Theoretical results are available on how good our approximation is.

Theorem 2.2.

(i) For a general partition of connected sets,

$$d_H(\mu, \mu_n) \leq \frac{1+s}{1-s} \max_{1 \leq i \leq n} \text{diam}(A_i),$$

(ii) If the partition is a regular cubic lattice, and a_i are chosen as center points of the partition sets,

$$d_H(\mu, \mu_n) \leq \frac{1+s}{2(1-s)} \max_{1 \leq i \leq n} \text{diam}(A_i),$$

Thus we have a rigorous upper bound for the distance between our approximation and the true invariant measure in terms of the natural metric for fractal measures.

Remark 2.3. In terms of computing time, one must calculate the images of rn points and for each of these points, a search must be performed over the n partition sets to find which set contains the image. At first glance, the computing time required by our method seems to be $O(n^2)$. However, by using the continuity of the maps T_k , the search may be restricted to only a few partition sets, since the images of neighboring points are close. Thus the construction of P_n takes $O(n)$ computing time.

Since unity is the largest eigenvalue of P_n , the power method may be used to find a fixed left eigenvector, and the computing time is negligible. The memory requirements for finding the left eigenvector are also $O(n)$ as the power method requires only the matrix P_n (which may be represented in sparse form, storing only the positions and values of nonzero entries) and the current test vector to be held in memory.

At this point we note that once we have calculated the matrices $P_n(k)$ for a given set of mappings T_1, \dots, T_r , it is a very simple matter to alter the probabilities w_k and recalculate an approximate invariant measure for this new system. This is because almost all of the computing effort goes into the construction of $P_n(k)$. For new probabilities w_k , we merely construct a new matrix P_n from (11) and find a fixed left eigenvector. Thus once we have an approximation for one set of probabilities, new approximations may be computed very quickly for a *whole range of probabilities*. The same idea applies to the estimation of Lyapunov exponents in the following sections and is briefly illustrated in the final example.

2.3. Example: The fern

We illustrate our results for a well-known IFS. Four affine mappings are used to produce a picture of a fern [Gutiérrez *et al.*, 1996]. In a later section, we will find the Lyapunov exponents of this IFS.

$$T_1x = \begin{pmatrix} 0.81 & 0.07 \\ -0.04 & 0.84 \end{pmatrix} x + \begin{pmatrix} 0.12 \\ 0.195 \end{pmatrix}$$

$$T_2x = \begin{pmatrix} 0.18 & -0.25 \\ 0.277 & 0.28 \end{pmatrix} x + \begin{pmatrix} 0.12 \\ 0.02 \end{pmatrix}$$

$$T_3x = \begin{pmatrix} 0.19 & 0.275 \\ 0.238 & -0.14 \end{pmatrix} x + \begin{pmatrix} 0.16 \\ 0.12 \end{pmatrix}$$

$$T_4x = \begin{pmatrix} 0.0235 & 0.087 \\ 0.045 & 0.1666 \end{pmatrix} x + \begin{pmatrix} 0.11 \\ 0 \end{pmatrix}.$$

The respective probabilities are $w_1 = 0.753$, $w_2 = 0.123$, $w_3 = 0.104$, $w_4 = 0.02$, while the contraction factors are $s_1 = 0.8480$, $s_2 = 0.3580$, $s_3 = 0.3361$, $s_4 = 0.1947$, yielding $s = 0.7215$. A representation of an approximation of the unique invariant measure using a 200×200 grid: $[0.005(g-1), 0.005g] \times [0.005(h-1), 0.005h]$, $g = 1, \dots, 200$, $h = 1, \dots, 200$ is shown in Fig. 1. As we choose grid center points $(0.0025(2g-1), 0.0025(2h-1))$, $g = 1, \dots, 200$, $h = 1, \dots, 200$, by Theorem 2.2, we know that

$$d_H(\mu_{40000}, \mu) \leq \frac{(1+0.7215)\sqrt{2}}{2(1-0.7215)} \frac{1}{200} \approx 0.0219.$$

To put this in perspective, if μ'_{40000} is the probability measure produced by translating the “picture” in Fig. 1 at a distance 0.0219 units, then $d(\mu_{40000}, \mu'_{40000}) \leq 0.0219$.

A direct comparison of these estimates with estimates of μ obtained via random iteration is difficult. One way to quantify the error is to run out a random orbit of length N and define μ_N as in (7). One may then calculate $d_H(\mu, \mu_N)$ and average this error over *all possible* random orbits. This would give an “expected error”, quantified in terms of the Hutchinson metric. For simple systems, such a theoretical analysis is possible, and in [Froyland, 1998] the one-dimensional Cantor system is considered, with our method clearly outperforming random iteration or the “chaos game”. For more complicated systems, our method has the advantage of possessing a rigorous upper bound on the error of the approximation; something which is lacking from approximations obtained from random iteration. On

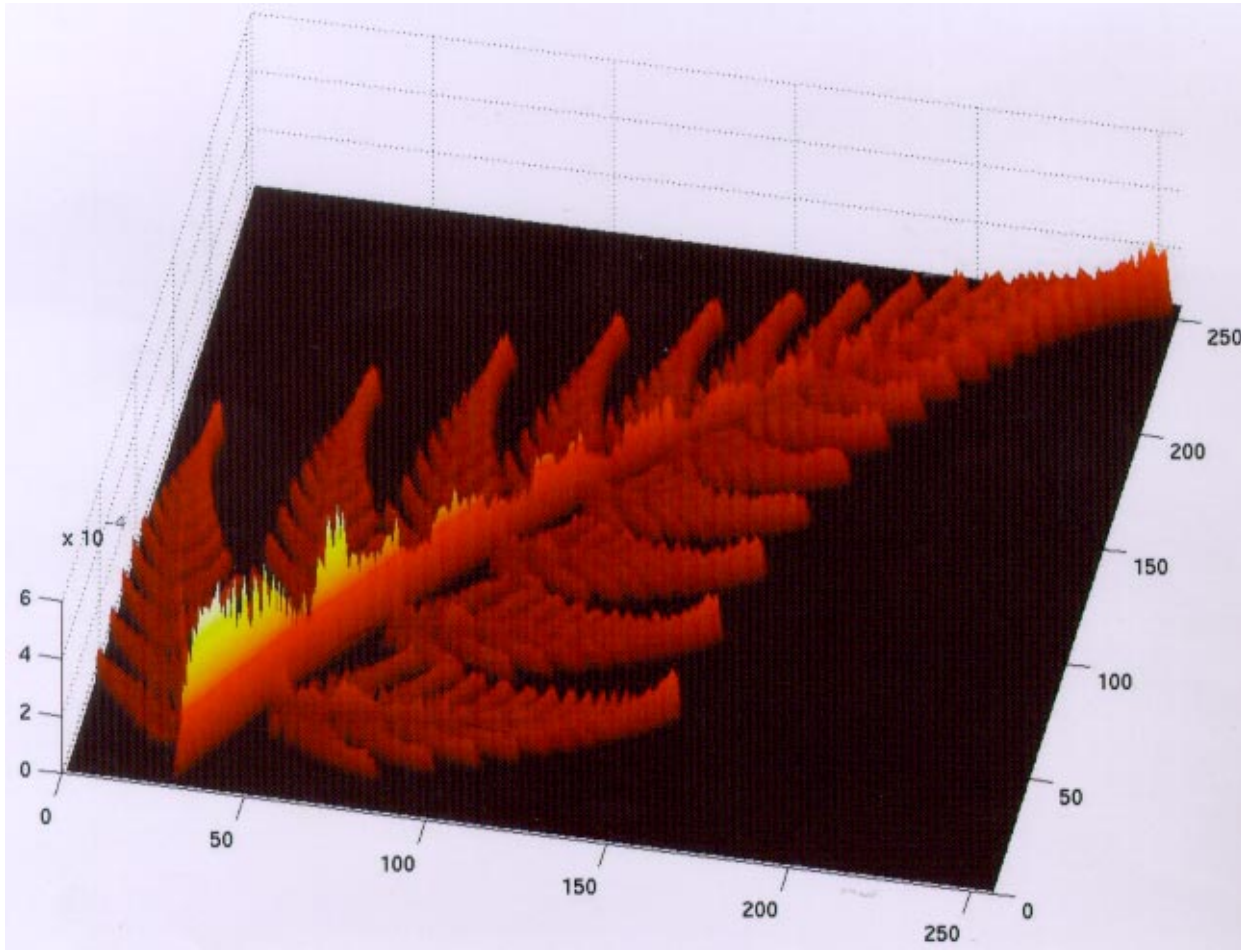


Fig. 1. A picture of a fern using a 200×200 grid. The density of μ_{40000} is plotted along the third coordinate axis (height).

an iteration for iteration basis in low-dimensional systems, we expect our discrete method to produce superior approximations of μ , with the rigorous bound of Theorem 2.2 often being very conservative. Intelligent ways of refining the partition (see [Dellnitz & Junge, 1998], for example) may also be used to increase the efficiency of the method.

3. Lyapunov Exponents of Iid Random Matrix Products

3.1. Background

We begin by discussing the Lyapunov exponents of an iid random matrix product (or affine IFS), as they are simpler to describe than the Lyapunov exponents of a nonlinear IFS. The latter will be dealt with in a later section.

An iid random matrix product is mathematically very similar to an IFS. Our IFS was defined by randomly selecting a map from some finite collection at each time step, and applying this selected map to the current point. By simply replacing the collection of maps $T_1 : M \rightarrow M, \dots, T_r : M \rightarrow M$ with a collection of $d \times d$ matrices $M_1 : \mathbb{R}^d \rightarrow \mathbb{R}^d, \dots, M_r : \mathbb{R}^d \rightarrow \mathbb{R}^d$, (treating them as maps on \mathbb{R}^d via matrix multiplication) we may now study the (random) evolution of an initial (nonzero) vector $v_0 \in \mathbb{R}^d$ under repeated multiplication by M_k , just as we studied the evolution of an initial point $x_0 \in M$ under the action of T_k . That is, we can define $v_N = v_N(k_{N-1}, \dots, k_0, v_0) := M_{k_{N-1}} \circ \dots \circ M_{k_0} v_0$ in analogy with (1), where composition now means multiplication. Because this system is simply linear (and not affine as in some IFS's), the dynamics is often rather boring, with v_i heading off towards infinity or contracting towards the origin. What is

of more interest is the *exponential rate* at which v_i diverges to infinity or converges to zero, and this rate is quantified by the Lyapunov exponents of the random matrix product. Precisely, one defines the limit:

$$\lambda^{(0)} = \lim_{N \rightarrow \infty} \log(\|M_{k_{N-1}} \circ \dots \circ M_{k_0} v_0\|^{1/N}). \quad (13)$$

Provided that

Hypotheses 3.1.

- (i) both $\|M_k\|$ and $\|M_k^{-1}\|$ are finite for $k = 1, \dots, r$, and
- (ii) there is no nontrivial subspace of \mathbb{R}^d that is left invariant by all M_k , $k = 1, \dots, r$,

the limit (13) exists for almost all random orbits. Moreover, when this limit exists, the same value of $\lambda^{(0)}$ is obtained for all such orbits and for all starting vectors v_0 ; see [Furstenberg & Kifer, 1983] for details. In the sequel, we will assume that Hypotheses 3.1 hold.³ Our matrices will be selected from $GL(d, \mathbb{R})$, the multiplicative group of invertible $d \times d$ real matrices. The invertibility condition ensures that our product does not suddenly collapse to zero if the current vector happens to be in the null space of the next matrix.

Before describing our method of calculation, we introduce a space that plays a central role in our construction.

Real projective space. It is clear from (13) that only the direction of v is important and not its length. We thus eliminate the unnecessary information and from now on consider our vectors to live in $d - 1$ -dimensional *real projective space* \mathbb{RP}^{d-1} (the “space of directions” in \mathbb{R}^d). This “space of directions” is the factor space $\mathbb{RP}^{d-1} = (\mathbb{R}^d \setminus \{0\}) / \sim$, where the equivalence relation \sim is defined by $u \sim v \Leftrightarrow u = a \cdot v$, for some $a \in \mathbb{R}$; $u, v \in \mathbb{R}^d \setminus \{0\}$. One can visualize \mathbb{RP}^{d-1} as one hemisphere of S^{d-1} , the $d - 1$ dimensional unit sphere. For example, \mathbb{RP}^1 may be considered to be the “right” half of S^1 in \mathbb{R}^2 (comprised of those points with non-negative x -coordinate) with the endpoints identified. That is, $\mathbb{RP}^1 = \{(\cos \theta, \sin \theta) \in \mathbb{R}^2 : -\pi/2 \leq \theta < \pi/2\}$. A natural mapping $\Pi : \mathbb{R}^d \setminus \{0\} \rightarrow \mathbb{RP}^{d-1}$ may be defined by extending a given vector $v \in \mathbb{R}^d$ in both directions to form a line passing through the origin

in the direction of v . This line intersects the sphere S^{d-1} in two antipodal points v^+ and v^- , with the former lying in the right hemisphere. We define $\Pi(v) = v^+$. The simplest way to describe a point $v^+ \in \mathbb{RP}^{d-1}$ is through angular coordinates, and in numerical calculations, we typically subdivide S^{d-1} using an equiangular grid. Invertible $d \times d$ matrices have a natural action on \mathbb{RP}^{d-1} under matrix multiplication. To compute the image of a point in \mathbb{RP}^{d-1} under a $d \times d$ matrix, one may convert the vector into Cartesian coordinates, multiply by the matrix, and then convert back to angular coordinates. From now on, we assume that this conversion process is done automatically, and we denote by v both a vector in $\mathbb{R}^d \setminus \{0\}$ and a point in \mathbb{RP}^{d-1} .

3.2. The Calculation of $\lambda^{(0)}$

We may rewrite (13) as

$$\lambda^{(0)} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\ell=0}^{N-1} \log \|M_{k_\ell} \bar{v}_\ell\|, \quad (14)$$

where \bar{v}_ℓ denotes a unit vector in the direction of v_ℓ . From this point onwards, vectors v_i and v will be assumed to have unit length. The sum (14) is a (random) *time average*, and may be converted into a *space average* [Furstenberg & Kifer, 1983],

$$\lambda^{(0)} = \sum_{k=1}^r w_k \int_{\mathbb{RP}^{d-1}} \log \|M_k v\| d\xi(v) \quad (15)$$

for some suitable probability measures ξ on \mathbb{RP}^{d-1} . In complete analogy to the IFS case, suitable probability measures satisfy

$$\xi = \sum_{k=1}^r w_k \xi \circ M_k^{-1}. \quad (16)$$

Following Sec. 2, we define an operator $\mathcal{D} : \mathcal{M}(\mathbb{RP}^{d-1}) \rightarrow \mathcal{M}(\mathbb{RP}^{d-1})$ by $\mathcal{D}\xi = \sum_{k=1}^r w_k \xi \circ M_k^{-1}$, and seek its fixed points. Unlike the operator \mathcal{P} in the IFS case, the operator \mathcal{D} is not a contraction with respect to the Hutchinson metric. This is because it is impossible for an invertible matrix M_k to be a local contraction at all $v \in \mathbb{RP}^{d-1}$, based on a simple conservation of mass argument (the image of \mathbb{RP}^{d-1} under some $M_k \in GL(d, \mathbb{R})$ is all of \mathbb{RP}^{d-1} ,

³Conditions (i) and (ii) are satisfied for “almost all” choices of matrices M_k , $k = 1, r$; see [Froyland, 1998] for a formal result.

while if M_k was a strict contraction everywhere, the area/volume of $M_k(\mathbb{R}P^{d-1})$ must be strictly less than that of $\mathbb{R}P^{d-1}$; a contradiction). Thus we are faced with the possibility of there existing more than one fixed point of \mathcal{D} . This does not overly concern us, as in most cases (and in particular, under Hypotheses 3.1), all fixed points of \mathcal{D} produce the same value $\lambda(0)$.

To approximate a fixed point of \mathcal{D} , we use exactly the same method as in Sec. 2. Namely, partition $\mathbb{R}P^{d-1}$ into m connected sets V_1, \dots, V_m of small diameter, choose a single point $v_i \in V_i$, $i = 1, \dots, m$, and for each matrix M_k , produce a matrix

$$D_{m,ij}(k) = \begin{cases} 1, & \text{if } M_k v_i \in V_j, \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

and combine them to form

$$D_m := \sum_{k=1}^r w_k D_m(k). \quad (18)$$

Let d_m denote a fixed left eigenvector of D_m (which necessarily has all elements non-negative, and has been normalized so that the sum of its elements is unity) and define an approximate fixed point of \mathcal{D} by

$$\xi_m = \sum_{i=1}^m d_{m,i} \delta_{v_i}. \quad (19)$$

We now have the following result regarding our estimate of $\lambda^{(0)}$.

Theorem 3.2. *Assume that Hypotheses 3.1 hold, and let $\{d_m\}_{m=m_0}^\infty$ be a sequence of eigenvectors as constructed above. Then*

$$\lambda_m := \sum_{k=1}^r w_k \sum_{i=1}^m d_{m,i} \log \|M_k v_i\| \rightarrow \lambda^{(0)} \quad \text{as } m \rightarrow \infty. \quad (20)$$

Remark 3.3.

- (i) We could of course, try to estimate the value of $\lambda^{(0)}$ directly from a random simulation of (13) for some large finite N . However, we present numerical evidence that this method is often inefficient and inaccurate, in particular in the cases where there are matrices that are near to singular, or matrices that are chosen very infrequently. For a theoretical analysis in a simple model, see [Froyland, 1998].

- (ii) As in Remark 2.3, the required computing time is $O(n)$.

3.3. Example: The fern

We return to the fern IFS, and proceed to compute its Lyapunov exponents. Since the Jacobian matrices of the four mappings T_1, \dots, T_4 are constant, the orbit of the IFS in M is unimportant; all that matters is the sequence in which the mappings are applied. Thus, the Lyapunov exponents of the fern IFS are the same as the Lyapunov exponents of an iid random product of the matrices

$$M_1 = \begin{pmatrix} 0.81 & 0.07 \\ -0.04 & 0.84 \end{pmatrix},$$

$$M_2 = \begin{pmatrix} 0.18 & -0.25 \\ 0.277 & 0.28 \end{pmatrix},$$

$$M_3 = \begin{pmatrix} 0.19 & 0.275 \\ 0.238 & -0.14 \end{pmatrix},$$

$$M_4 = \begin{pmatrix} 0.0235 & 0.087 \\ 0.045 & 0.1666 \end{pmatrix},$$

chosen with probabilities $w_1 = 0.753$, $w_2 = 0.123$, $w_3 = 0.104$ and $w_4 = 0.02$. Since the matrices M_1, \dots, M_4 satisfy Hypotheses 3.1, there is only one exponent that is almost always observed, and this is the exponent that we wish to estimate. Before we describe our results, we attempt to use the standard method of random iteration to estimate $\lambda^{(0)}$. We use (14) directly, by randomly selecting an initial vector v_0 , and then producing an iid sequence of matrices, selected from M_1, \dots, M_4 . At each iteration, (14) is applied to track the current estimate of $\lambda^{(0)}$; we truncate this process after $N = 10^6$ iterations. The results are shown in Fig. 2 for ten separate random orbits for $9 \times 10^5 \leq N \leq 10^6$. After 10^6 iterations there are still some minor fluctuations along individual orbits, but significantly different estimates are obtained from the ten sample orbits. Using the ten final estimates at $N = 10^6$, we calculate the mean and standard deviation as -0.443672 and 0.000654 , respectively; leaving uncertainty in the third decimal place, or a standard error of around 0.15%. In the sequel, we will use our method to produce an estimate that appears to be accurate up to five decimal places, with far fewer iterations required.

We now begin a description of our method. We seek to approximate a probability measure ξ on $\mathbb{R}P^1$

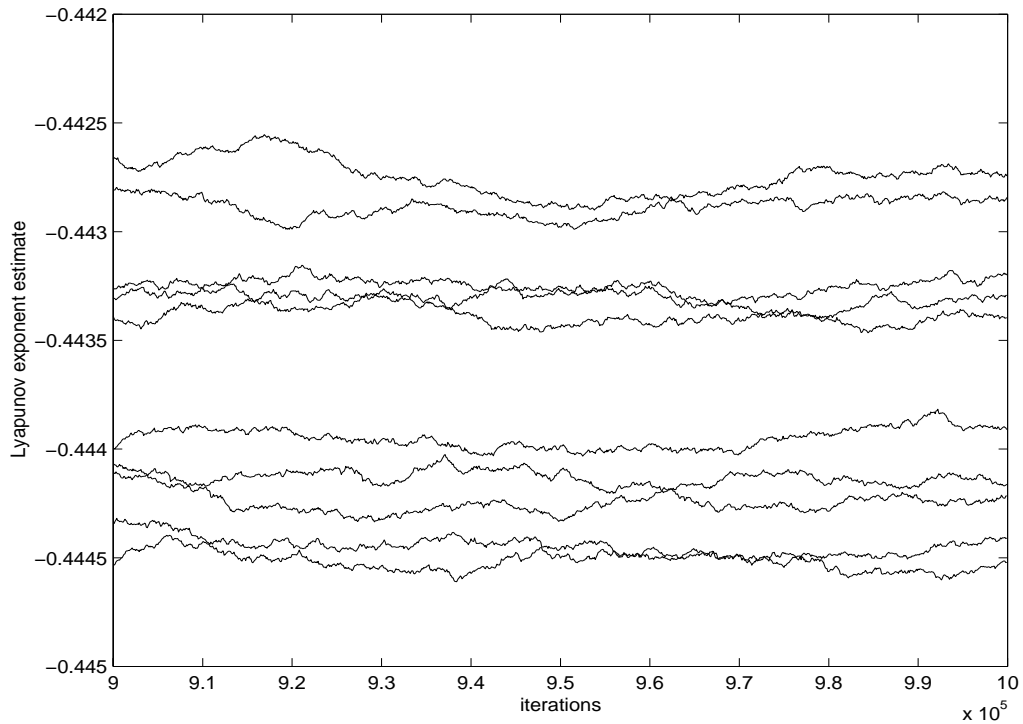


Fig. 2. Top Lyapunov exponent estimates calculated using (14) for ten random orbits of length 10^6 (shown between 9×10^5 and 10^6 iterations).

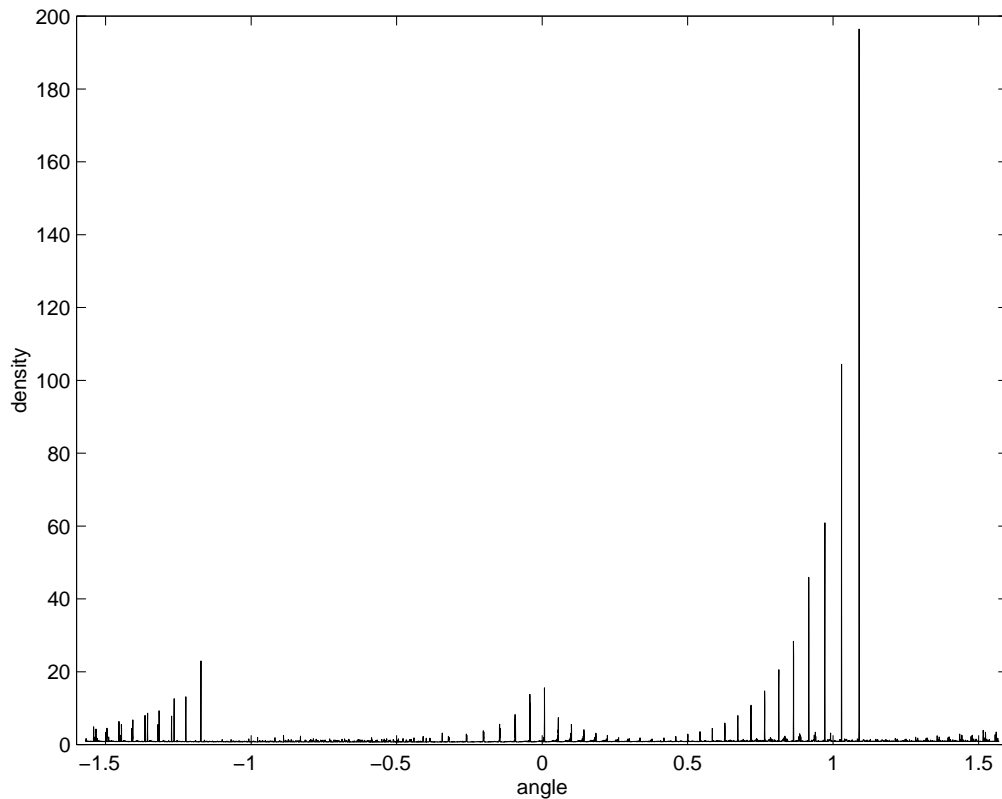


Fig. 3. Density of the approximate \mathcal{D} -invariant measure $\xi_{10\,000}$. This plot is a step function that takes the value $10\,000d_{10\,000,i}$ on the set $[-\pi/2 + (i-1)\pi/10\,000, -\pi/2 + i\pi/10\,000)$, $i = 1, \dots, 10\,000$.

that is invariant under \mathcal{D} . For numerical calculations, we treat $\mathbb{R}P^1$ as the subinterval $[-\pi/2, \pi/2)$. We construct equipartitions of $[-\pi/2, \pi/2)$ of cardinality $m = 100, 200, 1000, 2000, 10\,000, 20\,000, 40\,000$; these partitions consist of the sets $[-\pi/2 + (i - 1)\pi/m, -\pi/2 + i\pi/m)$, $i = 1, \dots, m$. The points v_i are chosen to be the center points of the partition sets, namely, $v_i = -\pi/2 + (2i - 1)\pi/2m$, $i = 1, \dots, m$. For a fixed partition (fixed value of m), we construct the matrix D_m from the four matrices $D_m(k)$, $k = 1, \dots, 4$. Producing each matrix $D_m(k)$ is a simple matter of applying the matrix M_k to each of the points v_i and noting which partition set contains the image. The matrices $D_m(k)$, $k = 1, \dots, 4$ are then added together to produce D_m , and the fixed left eigenvector d_m may be computed by a simple application of the power method (as the eigenvalue 1 is the largest eigenvalue of D_m). The density of the approximate \mathcal{D} -invariant probability measure $\xi_{10\,000}$ (from (19)) is shown in Fig. 3. By inserting the values of $d_{m,i}$ into (20), we produce an estimate of λ_m of the top Lyapunov exponent $\lambda^{(0)}$; see Table 1.

Remark 3.4. An alternative definition of the matrix $D_m(k)$ is $D'_{m,ij}(k) = \ell(V_i \cap M_k^{-1}(V_j)) / \ell(V_i)$, where ℓ is the normalized length on $[-\pi/2, \pi/2)$. This construction was proposed by Ulam [1964] for a related purpose. The length measure ℓ may be replaced by area and volume when working in $\mathbb{R}P^{d-1}$, $d = 2$ and 3 , respectively, however, the calculations become more difficult and Monte Carlo sampling methods or the like may be necessary to approximate the areas and volumes. In $\mathbb{R}P^1$, the construction of D'_m takes the same number of iterations (endpoints of the subintervals $[-\pi/2 + (i - 1)\pi/m, -\pi/2 + i\pi/m)$ are iterated, rather than the center points), and for

small m , better results are obtained. We denote the space average estimates produced by this alternative definition as λ'_m .

The results of Table 1 for low iteration numbers are plotted in Fig. 4. We compare our estimates with the time average estimate on an iteration for the iteration basis. The construction of each $D_m(k)$ requires m iterations of M_k , so D_m requires $4m$ iterations. We compare our estimate with the time average estimates after $N = 4m$ iterations; see Fig. 4(a). Figure 4(b) is a zoom of the area of Fig. 4(a) that contains the space average estimates. At the scale in (a), the two estimates λ_m and λ'_m coincide. Under the reasonable assumption that the estimates $\lambda_{40\,000}$ and $\lambda'_{40\,000}$ are accurate to five decimal places, we see that even with only 800 iterations, the estimates λ_{200} and λ'_{200} are at least as good as the estimate -0.443672 , which was obtained as a *mean* of ten time average estimates from trajectories of length 10^6 . Thus our method can produce more accurate estimates with far fewer iterations.

Remark 3.5. As is standard practice, the remaining exponents may be estimated by considering the action of the matrices on successively larger exterior powers of $\mathbb{R}P^{d-1}$ or $\mathbb{R}^d \setminus \{0\}$. Numerically, this means considering the action of the matrices on parallelepipeds rather than vectors. By considering the k th exterior powers (of k -dimensional parallelepipeds), the sum of the k largest exponents will be observed almost always, and our method is guaranteed to converge to this sum.

In particular, the sum of all of the exponents is easily calculated as

$$\begin{aligned} &\text{Sum of all Lyapunov exponents} \\ &= \sum_{k=1}^r w_k \log |\det M_k|. \end{aligned} \tag{21}$$

In our example, this value is -1.13005 , so based on Table 1, an estimate of the second exponent is $-1.13005 + 0.44351 = -0.68654$. The second (and smallest) exponent $\lambda^{(1)}$ may be calculated directly by substituting the inverses of M_1, \dots, M_4 in Eqs. (17) and (20). That is,

$$\lambda_m^- := - \sum_{k=1}^r \sum_{i=1}^m d_{m,i}^- \log \|M_k^{-1} v_i\| \rightarrow \lambda^{(1)} \quad \text{as } m \rightarrow \infty, \tag{22}$$

Table 1. Top Lyapunov exponent estimates for the fern IFS using a space average.

Number of Partition Sets m	Number of Iterations $N = 4m$	Estimate λ_m	“Ulam” Estimate λ'_m
100	400	-0.443078	-0.443535
200	800	-0.443508	-0.443457
1000	4000	-0.443556	-0.443494
2000	8000	-0.443481	-0.443505
10 000	40 000	-0.443506	-0.443510
20 000	80 000	-0.443513	-0.443511
40 000	160 000	-0.443509	-0.443509

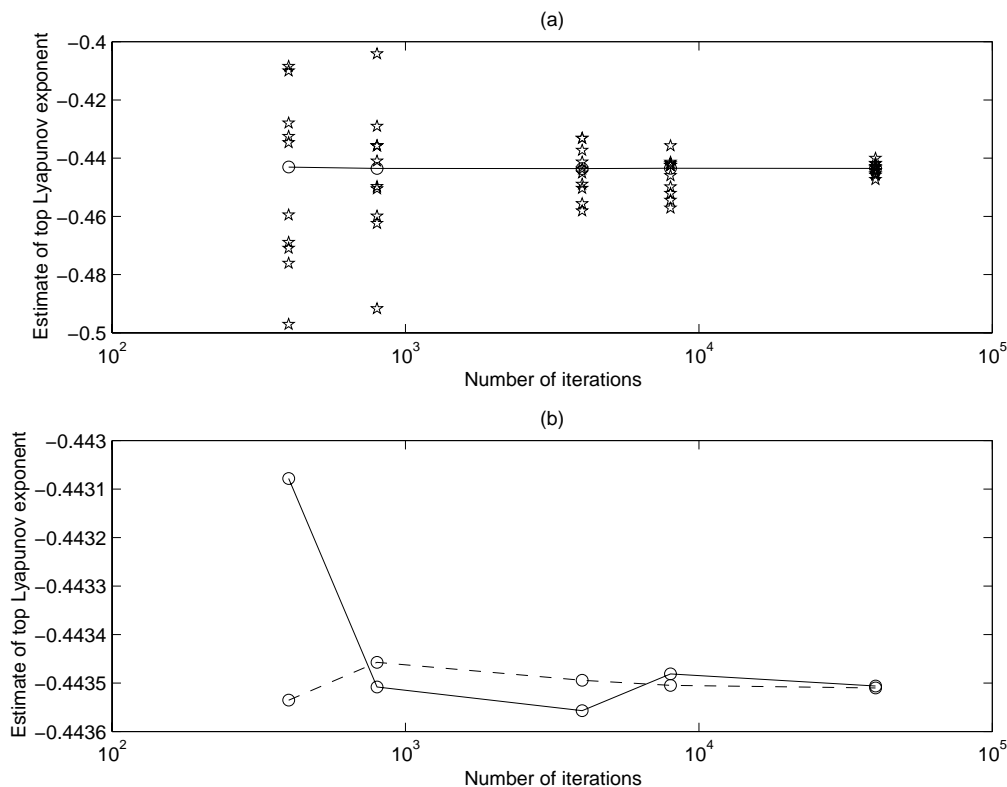


Fig. 4. (a) Plot of Lyapunov exponent estimates from ten random orbits versus number of iterations, $N = 400, 800, 4000, 8000, 40000$ (stars), and Lyapunov exponent estimates λ_m from the space average versus $N = 4m$ (circles). (b) Space average estimates λ_m (solid line) and λ'_m (dotted line) versus number of iterations $N = 4m$ (zoom of (a)).

where d_m^- is a fixed left eigenvector of the matrix D_m^- produced from (17) using M_k^{-1} in place of M_k . For $m = 10000$, we indeed obtain the value $\lambda_m^- = -0.68654$, establishing the consistency of our method.

4. Lyapunov Exponents of Iterated Function Systems

In this final section, we combine the methods of Secs. 2 and 3 to produce an approximation for the Lyapunov exponents of a nonlinear contractive IFS. The same technique may be applied to estimate the Lyapunov exponents of any iid random composition of maps, however in the noncontractive case, rigorous results are still lacking.

4.1. Background

We wish to estimate the values of λ that arise from the limit (2). Denote by μ the unique \mathcal{P} -invariant probability measure for our IFS.

Hypotheses 4.1.

- (i) $\int_M (\log^+ \|DT_k(x)\| + \log^+ \|DT_k(x)^{-1}\|) d\mu(x) < \infty$ for each $k = 1, \dots, r$.
- (ii) A “nondegeneracy” condition holds (see Theorem 4.10 in [Froyland, 1998] for a precise statement).

Under Hypotheses 4.1, only the top exponent $\lambda^{(0)}$ is observed for all vectors v and almost all random orbits of the IFS. In almost all cases, the nondegeneracy condition (ii) is satisfied (see discussion in [Froyland, 1998]).

4.2. Statement of results

As in Secs. 2 and 3, partition M into n connected sets A_1, \dots, A_n , and similarly, partition \mathbb{R}^{d-1} into m connected sets V_1, \dots, V_m . Choose a single point $a_i \in A_i$ for $i = 1, \dots, n$, and a single vector (angle coordinate) $v_g \in V_g$, for $g = 1, \dots, m$.

As in (11), we construct the matrices $P_n(k)$, $k = 1, \dots, r$ and find a normalized fixed left

eigenvector p_n , to approximate the unique invariant probability measure μ .

For each point $a_i \in M$, and map T_k , there is a Jacobian matrix $DT_k(a_i)$. We approximate the action of each $DT_k(a_i)$, $k = 1, \dots, r$, $i = 1, \dots, n$ using the technique described in Sec. 3. Define rn

$m \times m$ matrices

$$D_{m,gh}(k, i) = \begin{cases} 1, & \text{if } DT_k(a_i)v_g \in V_h, \\ 0, & \text{otherwise,} \end{cases} \quad (23)$$

$k = 1, \dots, r$, $i = 1, \dots, n$. Denoting $P_{n,ij}^*(k) = P_{n,ij}(k)p_{n,i}/p_{n,j}$, we now put the two approximations together to form the $nm \times nm$ matrix:

$$\mathcal{D}_{n,m} = \sum_{k=1}^r w_k \begin{pmatrix} P_{n,11}^*(k)D_m(k, 1) & P_{n,21}^*(k)D_m(k, 1) & \cdots & P_{n,m1}^*(k)D_m(k, 1) \\ P_{n,12}^*(k)D_m(k, 2) & P_{n,22}^*(k)D_m(k, 2) & \cdots & P_{n,m2}^*(k)D_m(k, 2) \\ \vdots & \vdots & \ddots & \vdots \\ P_{n,1m}^*(k)D_m(k, n) & P_{n,2m}^*(k)D_m(k, n) & \cdots & P_{n,mm}^*(k)D_m(k, n) \end{pmatrix} \quad (24)$$

Denote by $(s_m^{(1)}|s_m^{(2)}|\cdots|s_m^{(n)})$ a fixed left eigenvector of $\mathcal{D}_{n,m}$ that has been divided into segments of length m . Each segment is normalized so that $\sum_{g=1}^m s_{m,g}^{(i)} = 1$ for each $i = 1, \dots, n$.

Theorem 4.2. *Assume that Hypotheses 4.1 hold. Let $\{p_n\}_{n=n_0}^\infty$ be a sequence of normalized fixed left eigenvectors of P_n , and $\{(s_m^{(1)}|s_m^{(2)}|\cdots|s_m^{(n)})\}_{n=n_0, m=m_0}^\infty$ be a sequence of fixed left eigenvectors of $\mathcal{D}_{n,m}$, normalized as described above. Then*

$$\lambda_{n,m} := \sum_{k=1}^r w_k \sum_{i=1}^n p_{n,i} \times \sum_{g=1}^m s_{m,g}^{(i)} \log \|DT_k(a_i)v_g\| \rightarrow \lambda^{(0)} \quad \text{as } n, m \rightarrow \infty. \quad (25)$$

If our system is one-dimensional, our job becomes particularly easy, as we need only worry about approximating μ , and do not need to construct the matrix $\mathcal{D}_{n,m}$. Additionally, we can obtain *error bounds* in this situation.

Corollary 4.3 (One-Dimensional Version of Theorem 4.2). *Suppose that M is one-dimensional, and that $\int_M (\log^+ |T_k'(x)| + \log^+ |(T_k'(x))^{-1}|) d\mu(x) < \infty$ for each $k = 1, \dots, r$. Let $\varepsilon_n = \max_{1 \leq i \leq n} \text{diam}(A_i)$. In the notation of Theorem 4.2 define,*

$$\lambda_n := \sum_{k=1}^r w_k \sum_{i=1}^n p_{n,i} \log |T_k'(a_i)| \quad (26)$$

Then

$$|\lambda^{(0)} - \lambda_n| \leq \frac{2\varepsilon_n \max_{1 \leq k \leq r} \text{Lip}(\log |T_k'|)}{1 - s}.$$

Remarks 4.4.

- (i) The constructions of Theorem 4.2 and Corollary 4.3 may also be used to estimate the Lyapunov exponents of (noncontractive or chaotic) deterministic systems. In this case, there is only one mapping T_1 , which is applied with probability $w_1 = 1$. In such a situation, one is usually not guaranteed to have a unique invariant measure, however numerical results are encouraging; see [Froyland *et al.*, 1995] for an application of this technique using only time series data. It is often numerically observed that the Ulam construction of Remark 3.4 applied to the matrices P_n gives better results, especially for relatively small partitions.
- (ii) In (i) above, we are really approximating the action of a single mapping T_1 by the Markov chain governed by the transition matrix P . Thus, the same construction may be used to estimate the Lyapunov exponents of a Markovian product of matrices. That is, suppose there are n matrices M_1, \dots, M_n , and that the selection process is no longer iid, but Markovian, according to $\text{Prob}\{M_j \text{ is multiplied next} | M_i \text{ was just used}\} = P_{ij}$. A Markovian random matrix product can thus be defined and one often wishes to calculate the Lyapunov exponents of this product. In fact, $\lambda^{(0)}$ is given by (25),

putting $r = 1$, $w_1 = 1$, and

$$D_{m,gh}(i) = \begin{cases} 1, & \text{if } M_i v_g \in V_h, \\ 0, & \text{otherwise.} \end{cases} \quad (27)$$

This is illustrated in the final example.

4.3. Example: A random dimer model

This example demonstrates how to use (25) to compute the Lyapunov exponents of a Markovian random matrix product. The matrix product arises from a model known as the random dimer model [Oliveira & Petri, 1996] which we now briefly describe. Consider a discrete one-dimensional Schrödinger equation on an infinite lattice $\{\dots, -n, \dots, n, \dots\}$. At the n th lattice site, ψ_n denotes the amplitude of the wavefunction, and ε_n denotes the site energy. In the random dimer model, the site energy takes on only two values ε_a and ε_b . Furthermore, sites with energy ε_b can only occur in adjacent pairs; that is, a single site of energy ε_b may not be flanked on either side by sites of energy ε_a . The arrangement of the site energies on the lattice can be considered as a sample trajectory of a finite state Markov chain with the following transition matrix.

$$P = \begin{matrix} & \varepsilon_a & \varepsilon_b & \varepsilon_b \\ \varepsilon_a & \begin{pmatrix} p & 1-p & 0 \\ 0 & 0 & 1 \\ p & 1-p & 0 \end{pmatrix} \\ \varepsilon_b & & & \end{matrix}$$

This means, for example, that to the right of a site of energy ε_a , a site of energy ε_b occurs with probability $1 - p$, where $0 < p < 1$. Schrödinger's equation may be written in a recursive form [Oliveira & Petri, 1996; Crisanti *et al.*, 1993] as

$$\begin{pmatrix} \psi_{n+1} \\ \psi_n \end{pmatrix} = \begin{pmatrix} E - \varepsilon_n & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \psi_n \\ \psi_{n-1} \end{pmatrix}, \quad (28)$$

where E denotes the system energy. This recursive formula is nothing other than a Markovian product of matrices, drawn in this case from a set of two matrices (as ε_n takes only the two values ε_a and ε_b). An object of concern is the *localization length*, given by

$$L = - \lim_{|n| \rightarrow \infty} \frac{1}{|n|} \langle \log |\psi_n| \rangle$$

which measures how the wavefunction amplitudes vary with n . This length is simply the inverse of the top Lyapunov exponent of this random matrix product, which we now set out to estimate.

For the moment we shall fix $p = 1/2$. Setting $\varepsilon_a = \varepsilon_b = 0.5$ and $E = 0.7$, the three matrices which we will multiply together are

$$M_1 = \begin{pmatrix} 1.2 & -1 \\ 1 & 0 \end{pmatrix}, M_2 = M_3 = \begin{pmatrix} 0.2 & -1 \\ 1 & 0 \end{pmatrix}$$

Referring to Remark 4.4(ii), the transition matrix for this Markovian matrix product is given by

$$P = \begin{matrix} & 1 & 2 & 3 \\ 1 & \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \end{pmatrix} \\ 2 & & & \\ 3 & & & \end{matrix}$$

To begin, one forms the $n = 3$ matrices $D_{m,gh}(i)$, $i = 1, \dots, 3$ as in (27) (with $D_m(2) = D_m(3)$ since $M_2 = M_3$), and then constructs the $3m \times 3m$ matrix displayed in (24) (recalling that here $r = 1$ and $w_1 = 1$). The fixed left eigenvector s_m may be computed rapidly via the power method as unity is the largest eigenvalue of the (extremely sparse) matrix $\mathcal{D}_{3,m}$. Formula (25) may now be employed to produce an estimate $\lambda_{3,m}$ of the top Lyapunov exponent of this Markovian random matrix product.

Table 2. Lyapunov exponent estimates for the random dimer model using a space average.

# of Partition Sets m	# of Iterations $N = 2m$	Estimate/(CPU time) $\lambda_{3,m}$	"Ulam" Estimate/(CPU time) $\lambda'_{3,m}$
400	800	5.93373×10^{-3} (1.0s)	5.99273×10^{-3} (1.7s)
800	1600	5.96439×10^{-3} (3.0s)	5.99160×10^{-3} (5.3s)
1200	2400	5.98256×10^{-3} (6.1s)	5.99137×10^{-3} (10.9s)
1600	3200	5.99157×10^{-3} (10.3s)	5.99129×10^{-3} (18.6s)
2800	5600	5.98995×10^{-3} (57.5s)	5.99123×10^{-3} (89.7s)
4000	8000	5.99036×10^{-3} (116.7s)	5.99121×10^{-3} (184.7s)

This procedure was carried out for $m = 100, 200, 400, 600, \dots, 2000$, the results of which are displayed in Fig. 7 and in part in Table 2. The method in Remark 3.4 was also tested, and as before was found to give more stable results. Based on these data, it appears that the true value of the top Lyapunov exponent may be expressed to five significant digits as 5.9912×10^{-3} . Both the one-point and Ulam method have produced answers very close to this value using only around 3200 iterations, or a CPU time of 10–20 seconds, depending on which of the two methods is used.

In analogy to Fig. 3, we plot the densities on \mathbb{RP}^1 defined by the vectors $s_m^{(1)}$, $s_m^{(2)}$, and $s_m^{(3)}$ for $m = 1000$ in Fig. 5. These densities have the interpretation that if a uniform mass on \mathbb{RP}^1 is pushed forward along a Markov random trajectory of matrices (with the trajectory beginning in the infinite past), terminating at state i , then almost surely, the resulting distribution is approximated by $s_m^{(i)}$. The i th inner summation in (25) represents an

integration with respect to the i th density shown in Fig. 5.

We now compare these results with those obtained from (2) by simulating a long random orbit. In analogy to Fig. 2, Fig. 6 shows the computed estimates of the top Lyapunov exponent along the tail sections of ten orbits of length 10^6 . The mean and standard deviation of the values obtained at $N = 10^6$ are 5.95488×10^{-3} and 5.76197×10^{-5} , respectively, leaving an uncertainty in the second significant digit of the mean, or a standard error of around 0.97%. This mean and standard deviation is shown as a vertical bar on the right-hand side of Fig. 7. From the strong indications that the space average has converged to a value very close to the true value of $\lambda^{(0)}$, it appears that from ten random orbits of length 10^6 (taking a total CPU time of over 1900 seconds), the random iteration (or time average) method still produces comparatively inaccurate results.

Finally, we consider the variation of $\lambda^{(0)}$ as the transition probability p varies between 0 and 1. By

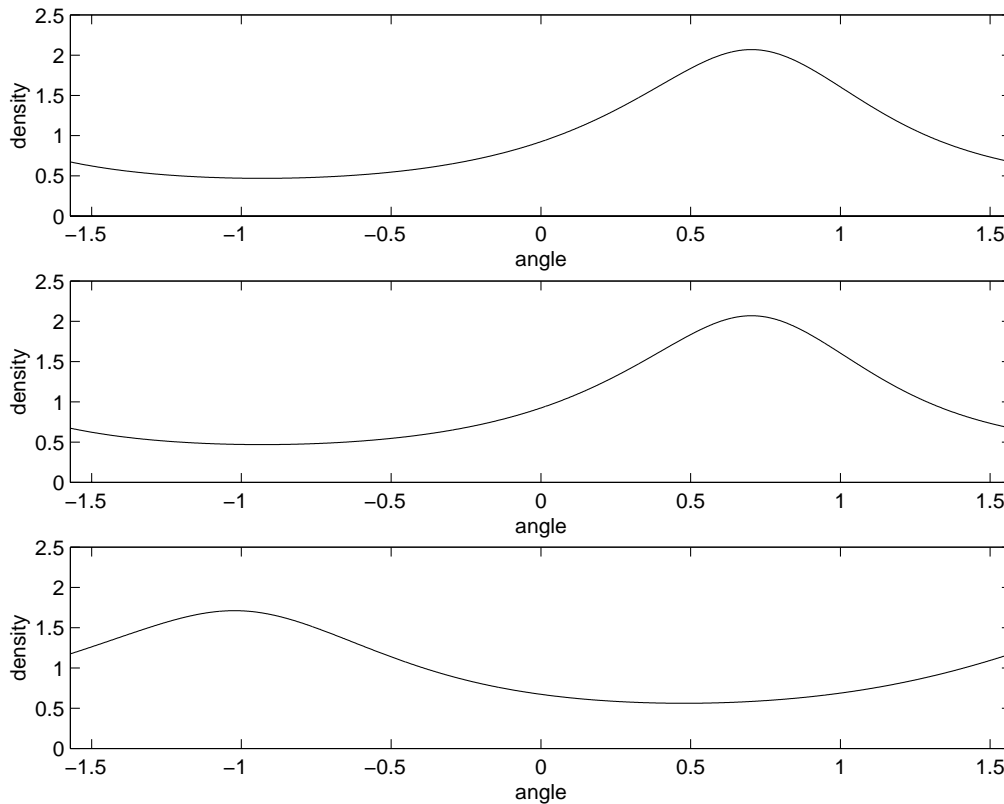


Fig. 5. Density of the vectors $s_{1000}^{(1)}$, $s_{1000}^{(2)}$, and $s_{1000}^{(3)}$ (from top to bottom, respectively). The interval $[-\pi/2, \pi/2)$ has been normalized to have length 1.

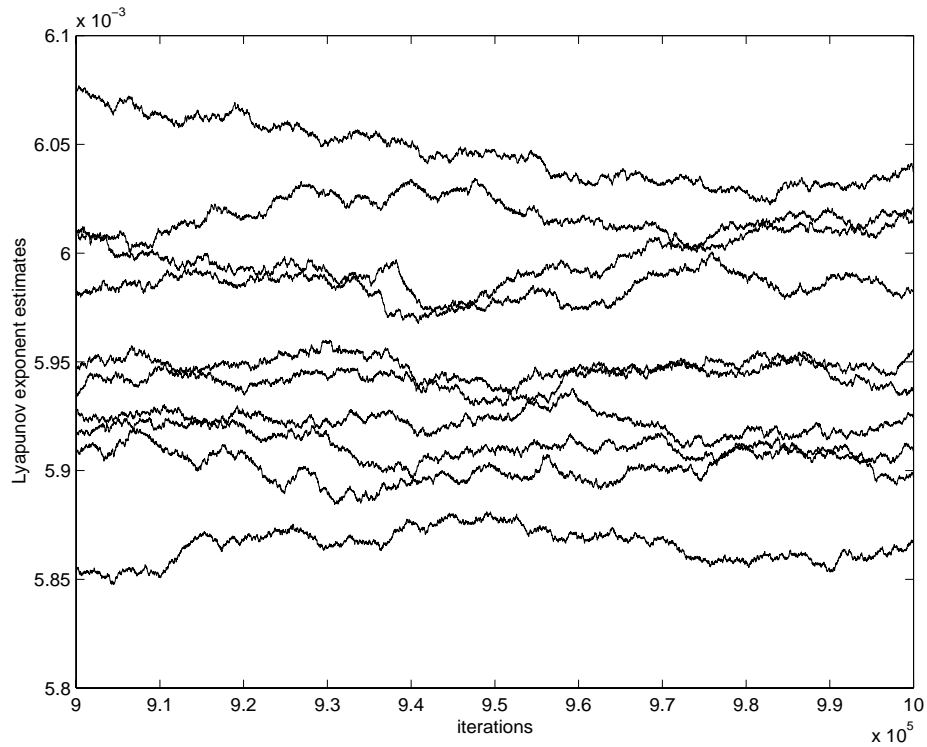


Fig. 6. Top Lyapunov exponent estimates for the random dimer model versus the number of iterations, calculated using (2) for ten simulated random orbits of length 10^6 (shown between 9×10^5 and 10^6 iterations).

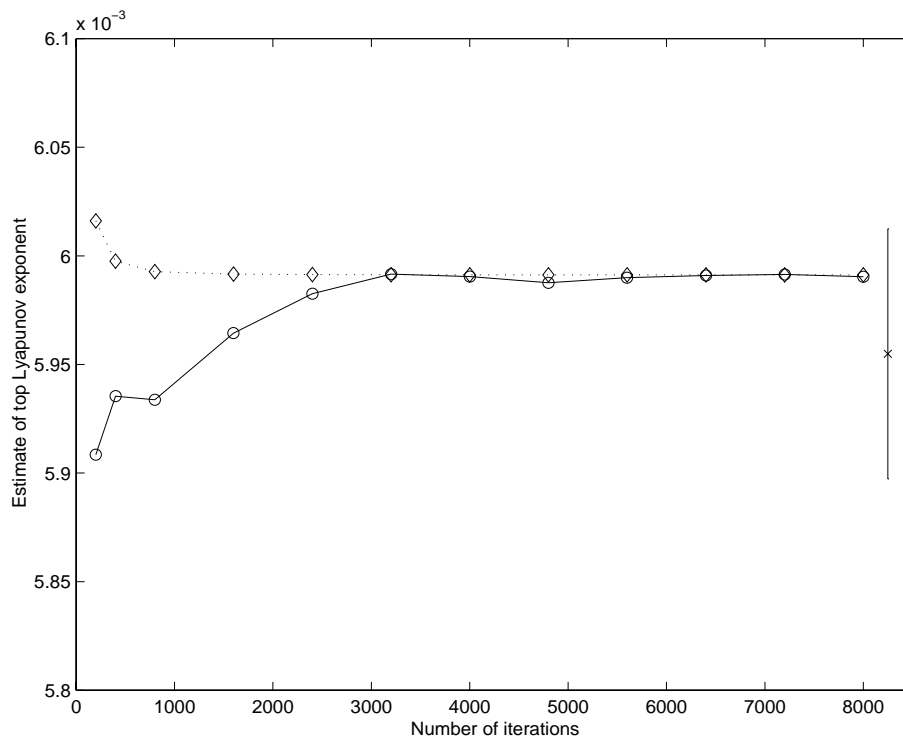


Fig. 7. Plot of space average estimates for the top Lyapunov exponent versus the number of iterations used. The solid line uses the one point method (27) of constructing the matrices $D_m(i)$ and the dotted line uses the alternative “Ulam” method outlined in Remark 3.4. The axes have been scaled to allow comparison with Fig. 6; the cross and vertical line to the right of the plot show the mean and one standard deviation (to each side of the mean) of the time average simulation in Fig. 6.

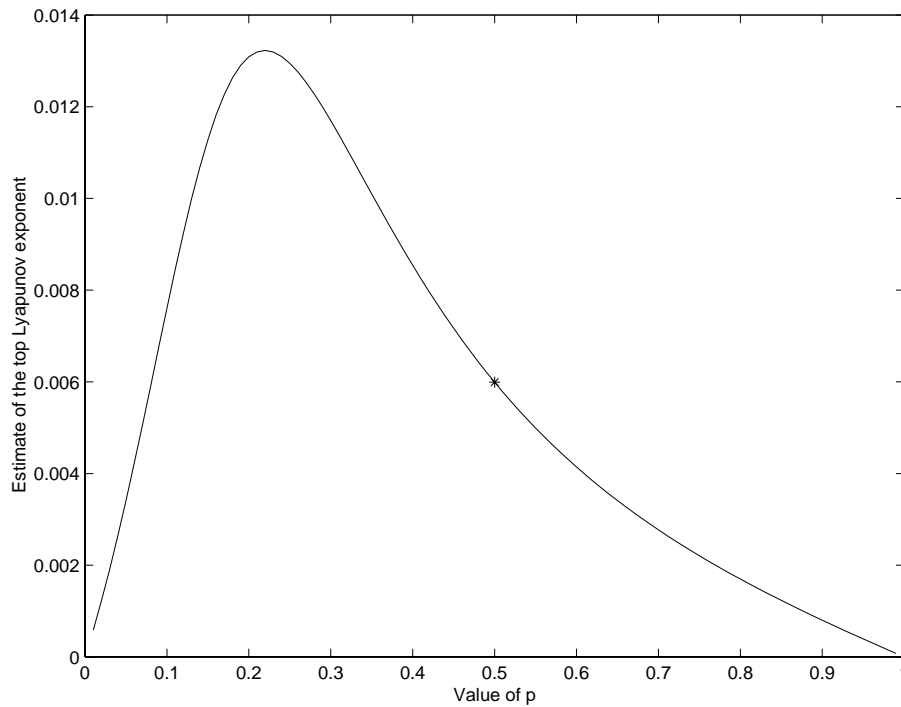


Fig. 8. Graph of $\lambda'_{3,m}(p)$ versus p , for p between 0 and 1 at 0.01 intervals, using $m = 1600$. Note that at $p = 0$ and $p = 1$, $\lambda^{(0)} = 0$ as $p = 0$ corresponds to only the matrices $M_2 = M_3$ occurring and $p = 1$ corresponds to only the matrix M_1 occurring (and each of the matrices have both eigenvalues on the unit circle). Note also that at $p = 0.5$ (indicated by the asterisk) we recover our earlier value for $\lambda^{(0)}$.

only altering the transition matrix P , we do not need to recalculate the matrices $D_m(i)$, the task which requires most of the computing effort. Thus, for each new P , we need only form the large matrix $\mathcal{D}_{3,m}$ in (24) (reusing the same $D_m(i)$), compute the fixed left eigenvector, and substitute the result into (25). The results are shown in Fig. 8. If one were to use the random iteration method, an *entirely new simulation* would be required for each value of p , making the task much more expensive.

Acknowledgments

The research by Gary Froyland was supported by the Japan Society for the Promotion of Science postdoctoral fellowship (research grant from the Ministry of Education, Science, and Culture No. 09-97004).

References

- Barnsley, M. F. & Elton, J. H. [1988] “A new class of Markov processes for image encoding,” *Adv. Appl. Probab.* **20**(1), 14–32.
- Benettin, G., Galgani, L., Giorgilli, A. & J.-M. Strelcyn [1980] “Lyapunov characteristic exponents for smooth

dynamical systems and for Hamiltonian systems; a method for computing all of them. Part 1: Theory,” *Meccanica* **15**, 9–20.

- Boyarsky, A. & Lou, Y. S. [1992] “A matrix method for approximating fractal measures,” *Int. J. Bifurcation and Chaos* **2**(1), 167–175.
- Crisanti, A., Vulpiani, A. & Paladin, G. [1993] *Products of Random Matrices in Statistical Physics*, Springer Series in Solid-State Sciences, Vol. 104 (Springer-Verlag, Berlin).
- Dellnitz, M. & Junge, O. [1998] “An adaptive subdivision technique for the approximation of attractors and invariant measures,” *Comput. Vis. Sci.* **1**, 63–68.
- de Oliveira, M. J. & Petri, A. [1996] “Generalized Lyapunov exponents for products of correlated random matrices,” *Phys. Rev.* **E53**(3), 2960–2963.
- Elton, J. [1987] “An ergodic theorem for iterated maps,” *Erg. Th. Dyn. Syst.* **7**, 481–488.
- Falconer, K. [1997] *Techniques in Fractal Geometry* (Wiley, Chichester).
- Froyland, G. “On estimating invariant measures and Lyapunov exponents for iid compositions of maps,” Unpublished notes. Available at <http://www.upb.de/math/~froyland>.
- Froyland, G., Judd, K. & Mees, A. I. [1995] “Estimation of Lyapunov exponents of dynamical systems using a spatial average,” *Phys. Rev.* **E51**(4), 2844–2855.

Furstenberg, H. & Kifer, Y. [1983] “Random matrix products and measures on projective spaces,” *Israel J. Math.* **46**(1&2), 12–32.

Gutiérrez, J. M., Iglesias, A. & Rodríguez, M. A. [1996] “A multifractal analysis of IFSP invariant measures with application to fractal image generation,” *Fractals* **4**(1), 17–27.

Hutchinson, J. E. [1981] “Fractals and self similarity,” *Indiana Univ. Math. J.* **30**, 713–747.

Lasota, A. & Mackey, M. C. [1994] *Chaos, Fractals, and Noise. Stochastic Aspects of Dynamics*, Applied Mathematical Sciences, Vol. 97, 2nd edition (Springer-Verlag, NY).

Peruggia, M. [1993] *Discrete Iterated Function Systems* (A K Peters, Wellesley).

Stark, J. [1991] “Iterated function systems as neural networks,” *Neural Networks* **4**, 679–690.

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

Appendix A Proofs

We give in detail the proof of Theorem 2.2. The

reader is guided through the proof of Theorem 3.2 and a sketch of the proof of Theorem 4.2 is given.

A.1. Invariant Measures: Proof of Theorem 2.2

There are three main steps in the proof. Firstly we show that \mathcal{P} is a contraction on the metric space $(\mathcal{M}(M), d_H)$ (Lemma A.1). We then use the matrix P_n to define an operator \mathcal{P}_n on $\mathcal{M}(M)$, and show that $\mathcal{P}\nu$ and $\mathcal{P}_n\nu$ are always close in the d_H metric (Lemma A.5). Finally, we combine these two properties in Lemma A.6 to provide a bound for $d_H(\mu, \mu_n)$, where μ_n is our approximate \mathcal{P} -invariant measure satisfying $\mu_n = \mathcal{P}_n\mu_n$.

Lemma A.1 (Contractivity). *Denote by s_k , the contraction constant of T_k , $k = 1, \dots, r$, and put $s = \sum_{k=1}^r w_k s_k$. Then,*

$$d_H(\mathcal{P}\nu_1, \mathcal{P}\nu_2) \leq s \cdot d_H(\nu_1, \nu_2). \tag{A.1}$$

Proof

$$\begin{aligned} d_H(\mathcal{P}\nu_1, \mathcal{P}\nu_2) &= \sup_{h \in \text{Lip}(1)} \left| \int_M \sum_{k=1}^r w_k h(T_k x) d\nu_1(x) - \int_M \sum_{k=1}^r w_k h(T_k x) d\nu_2(x) \right| \\ &\leq \sum_{k=1}^r w_k \sup_{h \in \text{Lip}(1)} \left| \int_M h \circ T_k d\nu_1 - \int_M h \circ T_k d\nu_2 \right| \\ &= \sum_{k=1}^r w_k \sup_{h \in \text{Lip}(s_k)} \left| \int_M h d\nu_1 - \int_M h d\nu_2 \right| \\ &= \sum_{k=1}^r w_k \sup_{h \in \text{Lip}(1)} \left| \int_M s_k h d\nu_1 - \int_M s_k h d\nu_2 \right| \\ &= \left(\sum_{k=1}^r w_k s_k \right) d_H(\nu_1, \nu_2) \quad \blacksquare \end{aligned}$$

Remark A.2. The result of Lemma A.1 was proven in [Hutchinson, 1987, Theorem 4.4.1], with $s = \max_{1 \leq k \leq r} s_k$.

We now use the matrices $P_n(k)$ to define a family of Markov operators that will approximate \mathcal{P} in the appropriate sense.

Definition A.3. Let $\nu \in \mathcal{M}(M)$ and define

$$\mathcal{P}_n\nu = \sum_{j=1}^n \left(\sum_{k=1}^n w_k \sum_{i=1}^n \nu(A_i) P_{n,ij}(k) \right) \delta_{a_j}. \tag{A.2}$$

In words, the action of this operator is: measure the set A_i , find which partition set A_j the point a_i is mapped into by T_k , and place a contribution of weight $w_k\nu(A_i)$ concentrated on the point a_j

(repeat for each $k = 1, \dots, r$). The utility of such a family of Markov operators is that (i) they have easily computable fixed points μ_n , and (ii) these \mathcal{P}_n -invariant measures μ_n are close to the true invariant measure μ .

Lemma A.4 (Invariance of μ_n). *The probability measure μ_n as defined in (12) is a fixed point of the Markov operator \mathcal{P}_n .*

Proof. Straightforward. ■

Lemma A.5 (Uniform Approximation). *Define $\varepsilon_n = \max_{1 \leq i \leq n} \text{diam}(A_i)$, and let s be defined as in Lemma A.6. Then*

$$\sup_{\nu \in \mathcal{M}(M)} d_H(\mathcal{P}\nu, \mathcal{P}_n\nu) \leq (1 + s)\varepsilon_n. \tag{A.3}$$

Proof. Choose any $\nu \in \mathcal{M}(M)$, and define $a_x = a_i$, where $x \in A_i$.

$$\begin{aligned} d_H(\mathcal{P}\nu, \mathcal{P}_n\nu) &= \sup_{h \in \text{Lip}(1)} \left| \int_M hd(\mathcal{P}\mu) - \int hd(\mathcal{P}_n\nu) \right| \\ &= \sup_{h \in \text{Lip}(1)} \left| \left(\int_M \sum_{k=1}^r w_k h(T_k x) d\nu(x) - \sum_{k=1}^r w_k \sum_{j=1}^n \left(\sum_{i=1}^n \nu(A_i) P_{n,ij}(k) \right) h(a_j) \right) \right| \\ &\leq \sum_{k=1}^r w_k \sup_{h \in \text{Lip}(1)} \left(\left| \int_M h(T_k x) d\nu(x) - \sum_{i=1}^n h(T_k a_i) \nu(A_i) \right| \right. \\ &\quad \left. + \left| \sum_{i=1}^n \left(h(T_k a_i) - \sum_{j=1}^n P_{n,ij}(k) h(a_j) \right) \nu(A_i) \right| \right) \\ &\leq \sum_{k=1}^r w_k \sup_{h \in \text{Lip}(1)} \left(\left| \sum_{i=1}^n \int_{A_i} (h(T_k x) - h(T_k a_i)) d\nu(x) \right| + \left| \sum_{i=1}^n (h(T_k a_i) - h(a_{j(i)})) \nu(A_i) \right| \right) \\ &\leq \sum_{k=1}^r w_k \sup_{h \in \text{Lip}(1)} \left(\sup_{x \in M} d(h(T_k x), h(T_k a_x)) + \max_{1 \leq i \leq n} d(h(T_k a_i), h(a_{j(i)})) \right) \\ &\leq \sum_{k=1}^r w_k \left(\sup_{x \in M} d(T_k x, T_k a_x) + \max_{1 \leq i \leq n} d(T_k a_i, a_{j(i)}) \right) \\ &\leq \sum_{k=1}^r w_k \left(s_k \sup_{x \in M} d(x, a_x) + \varepsilon_n \right) \\ &\leq \sum_{k=1}^r w_k (s_k + 1) \varepsilon_n. \quad \blacksquare \end{aligned}$$

Lemma A.6 (Error bound).

$$d_H(\mu, \mu_n) \leq \frac{(1 + s)\varepsilon_n}{1 - s}. \tag{A.4}$$

Proof

$$\begin{aligned} d_H(\mu, \mu_n) &= d_H(\mathcal{P}\mu, \mathcal{P}_n\mu_n) \\ &\leq d_H(\mathcal{P}\mu, \mathcal{P}\mu_n) + d_H(\mathcal{P}\mu_n, \mathcal{P}_n\mu_n) \\ &\leq s \cdot d_H(\mu, \mu_n) + \sup_{\nu \in \mathcal{M}(M)} d_H(\mathcal{P}\nu, \mathcal{P}_n\nu). \end{aligned}$$

A rearrangement provides the result. ■

**A.2. Lyapunov exponents of iid random matrix products:
Proof of Theorem 3.2**

We now replace M with \mathbb{RP}^{d-1} and $\mathcal{P} : \mathcal{M}(M) \curvearrowright$ with $\mathcal{D} : \mathcal{M}(\mathbb{RP}^{d-1}) \curvearrowright$. The main difference is that \mathcal{D} is not a contraction on $\mathcal{M}(\mathbb{RP}^{d-1})$, so we have no complete analogue of Lemma A.1. Instead, we replace contractivity of \mathcal{P} with continuity of \mathcal{D} . The proofs of the next three lemmas follow exactly as those for Lemmas A.1, A.4 and A.5 respectively.

Lemma A.7 (Continuity). *The operator \mathcal{D} is continuous; in particular, one has*

$$d_H(\mathcal{D}\zeta_1, \mathcal{D}\zeta_2) \leq \left(\sum_{k=1}^r w_k \ell_k \right) d_H(\zeta_1, \zeta_2),$$

where ℓ_k denotes the Lipschitz constant of $v \mapsto DT_k(v)$.

In analogy to Definition A.3,

Definition A.8. Define an approximate operator $\mathcal{D}_m : \mathcal{M}(\mathbb{RP}^{d-1}) \curvearrowright$ by

$$\mathcal{D}_m \zeta = \sum_{j=1}^m \left(\sum_{k=1}^r w_k \sum_{i=1}^m \zeta(V_i) D_{m,ij}(k) \right) \delta_{v_j}. \quad (\text{A.5})$$

As before, we have

Lemma A.9 (Invariance of ξ_m). *The probability measure*

$$\xi_m := \sum_{j=1}^m d_{m,j} \delta_{v_j}$$

is a fixed point of the Markov operator \mathcal{D}_m .

Lemma A.10 (Uniform Approximation). *Let $\varepsilon_m = \max_{1 \leq j \leq m} \text{diam}(V_j)$. Then*

$$\sup_{\zeta \in \mathcal{M}(\mathbb{RP}^{d-1})} d_H(\mathcal{D}_m \zeta, \mathcal{D}\zeta) \leq \left(1 + \sum_{k=1}^r w_k \ell_k \right) \varepsilon_m.$$

We now use the following general result.

Lemma A.11 (Invariance of limit). *Let (X, ϱ) be a compact metric space. Let $T : X \curvearrowright$ be continuous, and $T_n : X \curvearrowright, n = 1, 2, \dots$ be a family of maps such that*

$$\varrho(T_n x, T x) \rightarrow 0 \text{ uniformly as } n \rightarrow \infty. \quad (\text{A.6})$$

Suppose that each T_n has at least one fixed point x_n , and denote by \tilde{x} a limit of the sequence $\{x_n\}_{n=1}^\infty$. Then \tilde{x} is a fixed point of T .

Proof. Note that

$$\varrho(T\tilde{x}, \tilde{x}) \leq \varrho(T\tilde{x}, Tx_n) + \varrho(Tx_n, T_n x_n) + \varrho(x_n, \tilde{x}).$$

We may make the RHS as small as we like by choosing a suitably high value for n , since the first term goes to zero by continuity of T , the second term goes to zero uniformly by (A.6), and the third term goes to zero by hypothesis. ■

To apply this Lemma we set $X = \mathcal{M}(\mathbb{RP}^{d-1})$, and $\varrho = d_H$. Recall that the metric d_H generates the weak topology on $\mathcal{M}(\mathbb{RP}^{d-1})$, and that with respect to this topology, $\mathcal{M}(\mathbb{RP}^{d-1})$ is compact. We let the map T in Lemma A.11 be the operator $\mathcal{D} : \mathcal{M}(\mathbb{RP}^{d-1}) \curvearrowright$ defined in (A.5). By Lemma A.7, we have continuity of \mathcal{D} and by Lemma A.10 we see that $d_H(\mathcal{D}\zeta, \mathcal{D}_m \zeta) \rightarrow 0$ uniformly as $m \rightarrow \infty$. The measure ξ_m is fixed under \mathcal{D}_m by Lemma A.9, and applying Lemma A.11 we have that if $\tilde{\xi}$ is a limit of the sequence $\{\xi_m\}$, then $\tilde{\xi}$ is \mathcal{D} -invariant. We now simply insert ξ_m into Eq. (15) to obtain the result of Theorem 3.2. Hypothesis 3.1(ii) is needed so that any \mathcal{D} -invariant measure $\tilde{\xi}$ will produce the top Lyapunov exponent when inserted into Eq. (15); see [Furstenberg & Kifer, 1983].

**A.3. Lyapunov exponent of iterated functions systems:
Proof of Theorem 4.2 and Corollary 4.3**

In the case of nonlinear IFS's, we must consider probability measures on $M \times \mathbb{RP}^{d-1}$. The analogue of the operator \mathcal{D} is an operator $\mathcal{D}' : \mathcal{M}(M \times \mathbb{RP}^{d-1}) \curvearrowright$ given by

$$\mathcal{D}' \zeta' = \sum_{k=1}^r w_k \zeta' \circ \mathcal{T}_k^{-1} \quad (\text{A.7})$$

where $\mathcal{T}_k(x, v) = (T_k x, D_x T_k(v))$ for $(x, v) \in M \times \mathbb{RP}^{d-1}$. We wish to approximate fixed points ξ' of \mathcal{D}' and do so by finding fixed points of an operator $\mathcal{D}'_{m,n}$ constructed using the matrix $\mathcal{D}_{m,n}$ in (24). The construction of $\mathcal{D}'_{m,n}$ and the arguments showing that fixed points $\xi'_{m,n}$ converge to a fixed point of \mathcal{D}' follow exactly as in Sec. A.2. The nondegeneracy condition of Hypothesis 4.1(ii) is required so

that any \mathcal{D}' -invariant measure will provide us with the top Lyapunov exponent when inserted into an appropriate version of the space average formula.

This is just a sketch of the proof of Theorem 4.2; further technical details are given in [Froyland, 1998].

Proof of Corollary 4.3. Note that $\lambda^{(0)} = \sum_{k=1}^r w_k \int_M \log |T'_k| d\mu$, and that $\lambda_n = \sum_{k=1}^r w_k \int_M \log |\tilde{T}'_k| d\mu_n$, where $\tilde{T}'_k(x) = T'_k(a_i)$ for $x \in A_i$. Now,

$$\begin{aligned} \left| \int_M \log |T'_k| d\mu - \int_M \log |\tilde{T}'_k| d\mu_n \right| &\leq \left| \int_M \log |T'_k| d\mu - \int_M \log |T'_k| d\mu_n \right| + \int_M |\log |T'_k| - \log |\tilde{T}'_k|| d\mu_n \\ &\leq d_H(\mu, \mu_n) \text{Lip}(\log |T'_k|) + \text{Lip}(\log |T'_k|) \max_{1 \leq i \leq n} \text{diam}(A_i) \\ &\leq 2\varepsilon_n \text{Lip}(\log |T'_k|) / (1 - s), \end{aligned}$$

using the bound of Theorem 2.2(i), from which the result follows. ■