On the estimation of invariant measures and Lyapunov exponents arising from iid compositions of maps

Gary Froyland*
Department of Mathematical Engineering and Information Physics
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, JAPAN
Email: gary@sat.t.u-tokyo.ac.jp

Abstract

We present a method of approximating the unique invariant measure and associated Lyapunov exponents of a random dynamical system defined by the iid composition of a family of maps in situations where only one Lyapunov exponent is observed. As a corollary, our construction also provides a method of estimating the top Lyapunov exponent of an iid random matrix product. We develop rigorous numerical bounds for our invariant measure approximations and prove convergence of our Lyapunov exponent estimates to the true values. Comparisons between the invariant measure and Lyapunov exponents estimates generated by both our method and conventional random iteration are given.

Keywords: random dynamical system, iid composition, invariant measure, Lyapunov exponent
AMS Classification: Primary: 58F11, Secondary: 60G99, 65D99

1 Introduction

We begin by giving a brief synopsis of the contents of this paper. We have three main concerns.

Firstly, we consider the problem of rigorously approximating unique invariant measures of random dynamical systems defined by an iid composition of mappings $T_k : M \circlearrowleft$, $k = 1, \ldots, r$, where M is some compact subset of $\mathbb{R}^d$. In particular, we concentrate on systems which contract distance "on average" as it is well known that such systems possess a unique probability measure μ that is invariant under the expected action of the maps. It is this measure which we wish to approximate. We provide a simple and efficient means of producing numerical estimates of μ and derive error bounds for our estimates in terms of a natural metric that generates the weak topology on the space of Borel probability measures on M. Our first main result is Corollary 3.12; comparisons are made with the standard approximation technique via a numerical example.

*Current address: Department of Mathematics and Computer Science, University of Paderborn, Warburger Str. 100, D-33098 Paderborn, Germany. Email: froyland@math.uni-paderborn.de. URL: http://www-math.uni-paderborn.de/∼froyland

Secondly, if the maps T_k are differentiable, we may consider the Lyapunov exponents of this random dynamical system, as developed by Kifer [11] and others. We separate the case where the T_k are affine mappings as this situation is commonly considered in the iterated function systems (IFS) literature, and the corresponding theory is much simpler. The problem of estimating Lyapunov exponents of a random affine system is equivalent to estimating the Lyapunov exponents of an iid random matrix product, as studied by Furstenberg and Kifer [7] and many others. Under mild conditions on the collection of matrices, we develop a simple, rigorous numerical method to estimate the top Lyapunov exponent of such a random matrix product, and prove convergence to the true value. To achieve this, we explicitly approximate a probability measure on real projective space that is left invariant under the expected multiplicative action of the matrices. Except in the most trivial cases, the problem of rigorously calculating the Lyapunov exponents of random matrix products is still largely unsolved[1], and for low-dimensional systems, our method produces results that are superior to Monte Carlo trials. Our second main result is Proposition 4.9; comparisons with the standard estimation technique are carried out for two numerical examples.

Thirdly, we consider estimating Lyapunov exponents for the case of non-affine T_k, and again develop a rigorous method of calculating the top exponent for this random system, in situations where only one exponent is observed almost surely. To produce our estimate, we approximate both the unique invariant measure μ, and a family of probability measures on real projective space. Our third main result is Proposition 4.19. We conclude with a version of Proposition 4.19 for one-dimensional systems; in such a situation, our technique is particularly simple to implement.

For ease of presentation, the proofs are given for a finite famly of maps $\{T_k\}$, however if the mappings vary smoothly in the random parameter, our results may be easily extended to the case of uncountable families as is outlined in the third numerical example.

2 Definitions and Notation

Let (M, d) be a complete metric space and $\{T_k\}_{k=1}^r$ be a family of continuous mappings $T_k : M \circlearrowleft$ for each $k = 1, \ldots, r$. Denote by $\underset{\sim}{w}$ the $1 \times r$ probability vector $(w_1, \ldots, w_r)$; $\underset{\sim}{w}$ is an element of the hyperplane $W := \{\underset{\sim}{w} \in \mathbb{R}^r : w_1 + w_2 + \cdots + w_r = 1, w_i \geq 0, i = 1, \ldots, r\} \subset \mathbb{R}^r$.

Definition 2.1: A pair $(\underset{\sim}{w}, \{T_k\})$ defines a random dynamical system on M in the following way. Let $S = \{1, \ldots, r\}$ and define a probability measure ρ on S by $\rho(\{k\}) = w_k$. Let $\Omega = \prod_{i=0}^{\infty} S_i$, $S_i = S$, $i \geq 0$, and define the probability measure $\mathbb{P}$ on Ω by $\mathbb{P} = \prod_{i=0}^{\infty} \rho$. Further, define the skew product $\tau_1 : \Omega \times M \circlearrowleft$ by $\tau_1(\omega, x) = (\sigma\omega, T_{\omega_0} x)$, where $\sigma : \Omega \circlearrowleft$ is the left shift. Let $x \in M$; we define the N^{th} (random) iterate of x (denoted $x_N(\omega)$) by the second component of $\tau^N(\omega, x)$. That is, $x_N(\omega) = T_{\omega_{N-1}} \circ T_{\omega_{N-2}} \circ \cdots \circ T_{\omega_0} x$.

Remark 2.2: Such a random dynamical system is often called an *iterated function system*, or simply *IFS* for brevity.

A further assumption placed on our mappings $\{T_k\}$ is that our random system is a contraction on average.

[1] See "Question 3" [13] for a discussion.

Definition 2.3: Let s_k denote the Lipschitz constant of the mapping T_k;

$$s_k = \sup_{x,y \in M} \frac{d(T_k x, T_k y)}{d(x,y)}. \tag{1}$$

We call $s = \sum_{k=1}^{r} w_k s_k$ the *contraction constant* for the pair $(\underset{\sim}{w}, \{T_k\})$, and say that it is a *contraction on average* if $s < 1$.

3 Invariant Measures

Definition 3.1: Denote the space of all Borel probability measures on M with bounded support by $\mathcal{M}(M)$. A probability measure $\mu \in \mathcal{M}(M)$ will be called an *invariant measure of* $(\underset{\sim}{w}, \{T_k\})$ if μ is a fixed point of $\mathcal{P}$ (or $\mathcal{P}$-invariant), where $\mathcal{P} : \mathcal{M}(M) \circlearrowleft$ is defined by

$$\mathcal{P}\nu = \sum_{k=1}^{r} w_k \nu \circ T_k^{-1}, \quad \nu \in \mathcal{M}(M). \tag{2}$$

Clearly, fixed points of $\mathcal{P}$ need not be preserved by *any* of the maps T_k; what we are after is a measure that is preserved by the expected combined actions of the T_k.

Remark 3.2: The operator $\mathcal{P}$ is produced from arguments involving the deterministic representation of our random system as the skew product τ_1. Fixed points of $\mathcal{P}$ are a simple way of defining probability measures on $\Omega \times M$ that are invariant under the deterministic mapping τ_1. The notion of invariant measures for the skew product formalises what we mean for by invariant measures of our IFS; see Appendix A.

A good reason for choosing to define an invariant measure for $(\underset{\sim}{w}, \{T_k\})$ as one that is fixed by the operator $\mathcal{P}$ is the following. The (unique) Borel probability measure left invariant under $\mathcal{P}$ is exhibited by almost all random orbits of $(\underset{\sim}{w}, \{T_k\})$, for any starting point $x \in M$ [5]; (μ is "exhibited" by an orbit $\{x_0, x_1, \ldots\}$ means $\frac{1}{N}\sum_{i=0}^{N-1} \delta_{x_i} \to \mu$ weakly as $N \to \infty$).

Definition 3.3: Define a metric on $\mathcal{M}(M)$ by

$$d_H(\nu_1, \nu_2) = \sup_{h \in \mathrm{Lip}(1)} \left| \int_M h \, d\nu_1 - \int_M h \, d\nu_2 \right|, \tag{3}$$

where Lip(1) is the space of all Lipschitz functions $h : M \to \mathbb{R}$ with Lipschitz constant 1.

This metric is sometimes called the "Hutchinson metric" [9]; it is a complete metric on $\mathcal{M}(M)$. In the following subsections, we show that the Markov operator $\mathcal{P}$ is a contraction with respect to this metric. This contraction property tells us immediately that (i) $\mathcal{P}$ has a unique fixed point, and (ii) this unique fixed point may be approximated as a fixed point of a nearby Markov operator. We then define a simple numerical approximation of $\mathcal{P}$ and compute its fixed point. Using contraction mapping arguments, we provide a numerical bound for the accuracy of our approximation.

3.1 Contraction properties of $\mathcal{P}$

We begin by showing that $\mathcal{P}$ is a contraction on the metric space $(\mathcal{M}(M), d_H)$.

Lemma 3.4: *Denote by s_k, the contraction constant of T_k, $k = 1, \ldots, r$, and put $s = \sum_{k=1}^{r} w_k s_k$. Then,*

$$d_H(\mathcal{P}\nu_1, \mathcal{P}\nu_2) \leq s \cdot d_H(\nu_1, \nu_2), \tag{4}$$

PROOF:

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

□

Remark 3.5: The result of Lemma 3.4 was proven in Hutchinson [9] Theorem 4.4.1, with $s = \max_{1 \leq k \leq r} s_k$.

Remark 3.6: Barnsley and Elton [1] have proved existence and uniqueness of an invariant measure for such systems under weaker hypotheses, using more complicated machinery. As we are after concrete error bounds for our approximations, we do not strive for the most general conditions of existence and uniqueness.

Lemma 3.7: *Let $\mu_n \in \mathcal{M}(M)$ denote a probability measure left fixed by another Markov operator $\mathcal{P}_n$ that is close to $\mathcal{P}$. Let s be as in Lemma 3.4, and define*

$$\delta = \sup_{\nu \in \mathcal{M}(M)} d_H(\mathcal{P}\nu, \mathcal{P}_n\nu). \tag{5}$$

Then

$$d_H(\mu, \mu_n) \leq \frac{\delta}{1 - s} \tag{6}$$

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. □

3.2 A simple approximation and convergence results

For the remainder of the paper, M will denote a compact subset of $\mathbb{R}^d$. We define a crude, yet effective, approximation of $\mathcal{P}$ as follows. Construct a partition of M into n connected sets $\{A_1, \ldots, A_n\}$. From each set, choose a single point c_i, $i = 1, \ldots, n$ and for each mapping T_k define an $n \times n$ stochastic matrix $P_n(k)$ by setting

$$P_{n,ij}(k) = \begin{cases} 1, & \text{if } T_k c_i \in A_j, \\ 0, & \text{otherwise} \end{cases} . \tag{7}$$

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

Definition 3.8: Let $\nu \in \mathcal{M}(M)$ and define

$$\mathcal{P}_n \nu = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} w_k \sum_{i=1}^{n} \nu(A_i) P_{n,ij}(k) \right) \delta_{c_j}. \tag{8}$$

In words, the action of this operator is: measure the set A_i, find which partition set A_j the point c_i is mapped into by T_k, and place a contribution of weight $w_k \nu(A_i)$ concentrated on the point c_j (repeat for each $k = 1, \ldots, r$). 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 3.9: *Define the $n \times n$ matrix P_n as*

$$P_n := \sum_{k=1}^{r} w_k P_n(k) \tag{9}$$

and denote by p_n a fixed left eigenvector of P_n. Then the measure

$$\mu_n := \sum_{i=1}^{n} p_{n,i} \delta_{c_i} \tag{10}$$

is a fixed point of the Markov operator $\mathcal{P}_n$

PROOF: Straightforward. □

We do not concern ourselves with the uniqueness of the fixed left eigenvectors of P_n, as *all* such eigenvectors give rise to approximations of the unique invariant measure μ of our random system.

Remark 3.10: A simple and fast way to compute a fixed left vector p_n is to apply the power method; that is, select some starting vector (a vector of 1's, for example), and repeatedly multiply this vector by the matrix P_n until a fixed vector is approached to a desired accuracy. As unity is the largest eigenvalue of P_n, and there is usually a large spectral gap to the next eigenvalue, this method works well even for very large matrices. The matrix P_n is extremely sparse, and therefore only the non-zero entries and their positions need to be stored, providing a large saving on computer memory. Application of the power method requires only the sparse matrix P_n and the current vector to be stored; a total memory requirement of size $O(n)$.

We now derive a bound for δ in (5) in terms of the diameter of the partition sets and the contractivity constant s (see Definition 2.3).

Lemma 3.11: *Define $\epsilon_n = \max_{1\le i\le n} \operatorname{diam}(A_i)$, and let s be defined as in (3.7). Then*

$$\sup_{\nu\in\mathcal{M}(M)} d_H(\mathcal{P}\nu, \mathcal{P}_n\nu) \le (1+s)\epsilon_n. \tag{11}$$

PROOF: Choose any $\nu \in \mathcal{M}(M)$, and define $c_x = c_i$, where $x \in A_i$.

$$\begin{aligned}
d_H(\mathcal{P}\nu, \mathcal{P}_n\nu) &= \sup_{h\in\mathrm{Lip}(1)} \left| \int_M h\; d(\mathcal{P}\nu) - \int h\; d(\mathcal{P}_n\nu) \right| \\
&= \sup_{h\in\mathrm{Lip}(1)} \left| \left(\int_M \sum_{k=1}^{r} w_k h(T_k x)\; d\nu(x) - \sum_{k=1}^{r} w_k \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \nu(A_i) P_{n,ij}(k) \right) h(c_j) \right) \right| \\
&\le \sum_{k=1}^{r} w_k \sup_{h\in\mathrm{Lip}(1)} \left(\left| \int_M h(T_k x)\; d\nu(x) - \sum_{i=1}^{n} h(T_k c_i)\nu(A_i) \right| \right. \\
&\qquad \left. + \left| \sum_{i=1}^{n} \left(h(T_k c_i) - \sum_{j=1}^{n} P_{n,ij}(k) h(c_j) \right) \nu(A_i) \right| \right) \\
&\le \sum_{k=1}^{r} w_k \sup_{h\in\mathrm{Lip}(1)} \left(\left| \sum_{i=1}^{n} \int_{A_i} \left(h(T_k x) - h(T_k c_i) \right)\; d\nu(x) \right| \right. \\
&\qquad \left. + \left| \sum_{i=1}^{n} \left(h(T_k c_i) - h(c_{j(i)}) \right) \nu(A_i) \right| \right) \\
&\le \sum_{k=1}^{r} w_k \sup_{h\in\mathrm{Lip}(1)} \left(\sup_{x\in M} d(h(T_k x), h(T_k c_x)) + \max_{1\le i\le n} d(h(T_k c_i), h(c_{j(i)})) \right) \\
&\le \sum_{k=1}^{r} w_k \left(\sup_{x\in M} d(T_k x, T_k c_x) + \max_{1\le i\le n} d(T_k c_i, c_{j(i)}) \right) \\
&\le \sum_{k=1}^{r} w_k \left(s_k \sup_{x\in M} d(x, c_x) + \epsilon_n \right) \\
&\le \sum_{k=1}^{r} w_k \left(s_k + 1 \right) \epsilon_n.
\end{aligned}$$

□

Thus, we have convergence of the easily computable μ_n to the invariant measure μ, with an explicit error bound depending on the maximal diameter of the partition sets.

Corollary 3.12 (Main Result 1): *Let μ_n be defined as in (10), and ϵ_n and s be as in Lemma 3.11. Then*

$$d_H(\mu, \mu_n) \le \frac{\epsilon_n(1+s)}{1-s} \tag{12}$$

PROOF: Immediate from Lemmas and 3.7 and 3.11. □

Remark 3.13: The book of Peruggia [14] has related constructions to ours to approximate μ in a similar setting. However, there are no results concerning the accuracy of these approximations, and his method of proving convergence of his approximations relies on the more technical work of [5].

Remark 3.14: After completion of the material in §3.2, the paper of Stark [15] came to our attention, in which similar approximation methods are detailed with a view to implementation on a neural network.

3.3 An Analytic Comparison: Expected errors in random iteration

In this section we compare our method with the standard method of random iteration via an example which lends itself to analytic methods. Consider the following situation. We have a random dynamical system, and decide that a good way to obtain an approximation of the unique invariant measure is to select a random (with respect to the unique invariant measure) starting point, and run out a random orbit of N points, placing a δ-measure of height $1/N$ at each of the points along the orbit. There will of course be some error in this approximation, which may be quantified by the Hutchinson metric. We can now ask "What is the expected error?", where we average the errors over all possible orbits (and all possible starting points). And perhaps more importantly "At what rate does the expected error decrease to zero as $N \to \infty$?". Unfortunately, the author knows of no general solution to such questions. The result of Elton [5], for example, states only that weak convergence occurs for all starting points and almost all random orbits, but orders of convergence are not treated.

Here we consider such questions for a simple well-known IFS. Consider the mappings $T_1, T_2 : [0,1] \circlearrowleft$, defined by $T_1 x = x/3$ and $T_2 x = x/3 + 2/3$. Our random dynamical system will be an iid composition of these mappings where each of the two maps has probability $1/2$ of being chosen at each time step. The maximal invariant set of this IFS is the middle thirds Cantor set $\Lambda = \bigcap_{i=0}^{\infty} (T_1^i[0,1] \cup T_2^i[0,1])$. We name our IFS the "Cantor system", for want of a better name. To define the unique invariant measure μ for this IFS is a little technical; as we do not know of a reference, its construction and proof of invariance are given in Appendix B. For our current purposes, we only need the facts that the subinterval $[0,1/3]$ is measurable and that $\mu([0,1/3]) = 1/2$.

3.3.1 The random iteration method

A common approximation of μ is constructed by running out long random orbits as described above, denoting the measure consisting of a convex combination of N δ-measures by μ_N^R (the superscript for "R"andom). In order to evaluate $d_H(\mu, \mu_N^R)$, we must integrate both μ and μ_N^R with respect to some Lipschitz function h as in (3). We shall choose a *single function* $C : [0,1] \to \mathbb{R}$ for the integrals in (3); the value obtained will be a *lower bound* for $d_H(\mu, \mu_N^R)$. It is desirable for the chosen function C to have the following properties:

(i) C should "distinguish" the two measures μ and μ_N^R in the sense that $\int_{[0,1]} C \, d\mu$ and $\int_{[0,1]} C d\mu_N^R$ differ as much as possible. One way this could be achieved is by $C(x)$ taking on large values near high density regions of μ, and taking on small values near high density areas of μ_N^R.

(ii) In order to maximise the lower bound for $d_H(\mu, \mu_N^R)$, C should have a low Lipschitz constant, since for general Lipschitz functions $h : M \to \mathbb{R}$,

$$d_H(\mu, \nu) = \sup_{h \text{ is Lipschitz}} \frac{1}{\mathrm{Lip}(h)} \left| \int_M h \, d\mu - \int_M h \, d\nu \right|. \tag{13}$$

Properties (i) and (ii) will of course have to be traded off against one another.

(iii) C must be simple enough so that it is possible to analytically average

$$\left| \int_{[0,1]} C \, d\mu - \int_{[0,1]} C \, d\mu_N^R \right|$$

over all possible random orbits of the Cantor system (where each orbit produces a different μ_N^R). This average will provide us with a lower bound for the *expected* difference between μ and μ_N^R.

Of these properties, property (iii) is of greatest importance. We therefore select a function that fulfils property (iii), but is probably far from being optimal for properties (i) and (ii). This is OK; it just means that our lower bound will be rather conservative. The function we choose is $C : [0, 1] \to \mathbb{R}$ defined by:

$$C(x) = \begin{cases} 1, & 0 \le x \le 1/3, \\ 2 - 3x, & 1/3 < x \le 2/3, \\ 0, & 2/3 < x \le 1. \end{cases} \tag{14}$$

The graph of C is shown in Figure 1. Since the supports of μ and μ_N^R are contained in

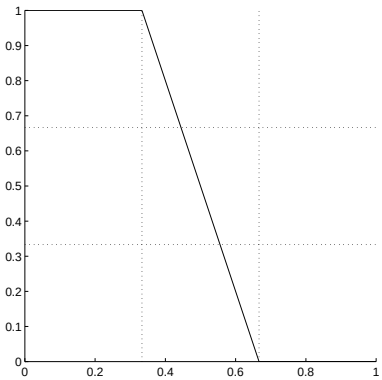

Figure 1: The graph of the function $C(x)$ defined in (14).

$[0, 1/3] \cup [2/3, 1]$, it is clear that

$$\int_{[0,1]} C \, d\mu = \mu([0, 1/3]) \quad \text{and} \quad \int_{[0,1]} C \, d\mu_N^R = \mu_N^R([0, 1/3]). \tag{15}$$

Thus we have reduced the difference of the integrals of C to the difference in the quantities $\mu([0, 1/3])$ and $\mu_N^R([0, 1/3])$.

Let the sequence $x_0, x_1, \ldots, x_{N-1}$ be a realisation of an orbit of the Cantor system. We are essentially interested in the value of $\mu_N^R([0, 1/3]) = \frac{1}{N} \sum_{i=0}^{N-1} \chi_{[0,1/3]}(x_i)$. Because of the special structure of the Cantor system, the function

$$\chi_{[0,1/3]}(T_{\omega_{N-2}} \circ \cdots T_{\omega_1} \circ T_{\omega_0}(x)) = \begin{cases} 1, & \text{if } \omega_{N-2} = 0, \\ 0, & \text{if } \omega_{N-2} = 2. \end{cases}$$

while for the $N = 0$ case, clearly

$$\chi_{[0,1/3]}(x) = \begin{cases} 1, & \text{if } x \in [0, 1/3], \\ 0 & \text{if } x \in [2/3, 1]. \end{cases}$$

Thus, the functions $H_i(\omega) := \chi_{[0,1/3]}(T_{\omega_{i-1}} \circ \cdots T_{\omega_1} \circ T_{\omega_0}(x))$, $i = 0, \ldots, N-1$ are Bernoulli random variables with probability $1/2$ of taking on either value 0 or 1. The expected value of H_i is $1/2$, and the variance is $1/4$. By setting $S_N = \frac{1}{N}\sum_{i=0}^{N-1} H_i$, we obtain a random variable whose values are the estimates $\mu_N^R([0, 1/3])$ produced by single realisations of the Cantor system.

The random variable S_N has the $(1/2, 1/2)$ N-trial binomial distribution. As $\mu([0, 1/3]) = 1/2$, we are interested in the mean deviance of S_N from $1/2$, and so introduce the random variable D_N which takes the value $|1/2 - i/N|$ with probability $\binom{N}{i}/2^N$. The mean of D_N may be computed as

$$\mathbb{E}(D_N) = \sum_{i=0}^{N} \left|\frac{1}{2} - \frac{i}{N}\right| \binom{N}{i} \left(\frac{1}{2}\right)^N = \begin{cases} \binom{N-1}{(N-1)/2}/2^N, & \text{for odd } N, \\ \binom{N}{N/2}/2^{N+1}, & \text{for even } N. \end{cases}$$

Using error bounds for Stirling's approximation of the binomial coefficients ([2], p.45), we find that

$$\frac{1}{\sqrt{2\pi N}} e^{-5/18N} < \mathbb{E}(D_N) < \frac{1}{\sqrt{2\pi N}} e^{-17/84(N-1)} \left(N/(N-1)\right)^{1/2} \tag{16}$$

Now, since C has a Lipschitz constant of 3, one has from (13):

$$\begin{aligned} d_H(\mu, \mu_N^R) &\geq \frac{1}{3}\left|\int_{[0,1]} C(x)\, d\mu(x) - \int_{[0,1]} C(x)\, d\mu_N^R(x)\right| \\ &= \left|\mu([0, 1/3]) - \mu_N^R([0, 1/3])\right|/3. \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{E}\left(d_H(\mu, \mu_N^R)\right) &\geq \mathbb{E}\left(\left|\mu([0, 1/3]) - \mu_N^R([0, 1/3])\right|\right)/3 \\ &= \mathbb{E}(D_N)/3 \\ &> \frac{1}{3\sqrt{2\pi N}} e^{-5/18N} \end{aligned}$$

by the LHS of (16). The bounds of (16) tell us that by using the approximation $\mathbb{E}(D_N) \gtrsim 1/\sqrt{2\pi N}$, we have *relative* errors of around 0.3% at $N = 100$, and 0.03% at $N = 1000$. The upshot of this is that we lose very little accuracy if we say that the lower bound for $\mathbb{E}\left(d_H(\mu, \mu_N^R)\right)$ is $1/3\sqrt{2\pi N}$, and in particular, that this expected error is of order $O(1/\sqrt{N})$.

3.3.2 Our method

Let us now compare these results with a bound for the distance between μ and and a fixed point of our perturbed Markov operator. We will make the comparison on an iteration for iteration basis. To construct the approximation $\mu_N^R([0, 1/3])$, we used N iterations of our map (counting the placement of the initial point x as one iteration for brevity). We ignore the generation of the $N - 1$ random numbers that were required to compute the random orbit, and will allow ourselves only N iterations to construct our approximation. Since we

are required to construct two transition matrices for each of the two maps T_1, T_2, we can only use $N/2$ iterations for each matrix. This amounts to constructing $P_{N/2}(1)$ and $P_{N/2}(2)$ using a grid of $N/2$ points each. By equidistributing these points in the unit interval, we may produce a value[2] of $1/N$ for $\epsilon_{N/2}$. Putting $s = (1 + 1/3)/(1 - 1/3) = 2$, Corollary 3.12 then gives us the bound

$$d_H(\mu, \mu_{N/2}) \leq 2(1/N) = 2/N. \tag{17}$$

The important thing to note is that our estimates converge in the d_H metric at a rate of $1/N$.

Clearly our bound is very conservative; in numerical simulations, our method performed much better than these bounds suggested. Nevertheless, for means of comparison with the time-average estimate, we use the bound of $2/N$. The first plot in Figure 2 shows the lower bound for $\mathbb{E}\left(d_H(\mu, \mu_N^R)\right)$ (dotted), and the upper bound for $d_H(\mu, \mu_{N/2})$ (solid). The difference between $O(1/\sqrt{N})$ and $O(1/N)$ is apparent with our method outperforming

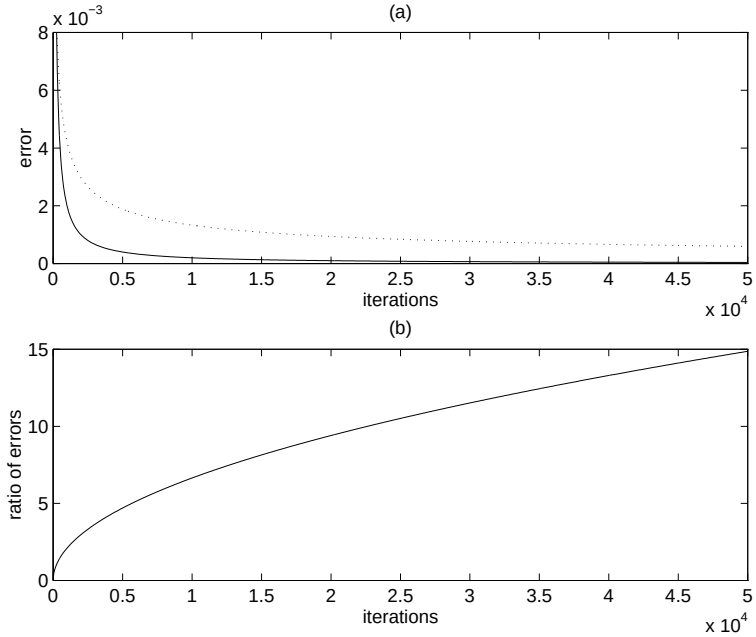

Figure 2: (a) A plot of $e^{-5/18N}/3\sqrt{2\pi N}$ vs N (dotted), a lower bound (16) for $\mathbb{E}\left(d_H(\mu, \mu_N^R)\right)$; and a plot of $2/N$ vs N (solid), the upper bound (17) for $d_H(\mu, \mu_{N/2})$. (b) A plot of $\left(e^{-5/18N}/3\sqrt{2\pi N}\right)/(2/N)$ vs N, the ratio: lower bound of $\mathbb{E}\left(d_H(\mu, \mu_N^R)\right)$/upper bound of $d_H(\mu, \mu_{N/2})$.

the time average after around 250 iterations. The second plot shows the ratio of the time average error to the error in our method; at 1000 iterations, our method is already producing estimates that are more than twice as accurate.

Remarks 3.15: Of course, in higher dimensional examples, we do not expect the difference in the two methods to be so pronounced. But, for low dimensional systems, possibly utilising intelligent means of partitioning the state space (as in Dellnitz [4], for example), we expect our method to remain computationally superior to random iteration.

[2]Since we choose the c_i to be the midpoints of sets of width $2/N$, the distance $d(T_k c_i, c_{j(i)})$ in the proof of Lemma 3.11 is $\leq 1/N$.

A further advantage of our method is that once the matrices $P_n(k)$ have been constructed for a given set of mappings $T_1, \ldots, T_r$, one may vary the probabilities w_k and compute invariant measures for these new systems very quickly. This is because almost all of the computing effort goes into the construction of the $P_n(k)$, and by altering the probabilities with which the maps are applied, we need only form a new P_n as in (7) and compute a new fixed left eigenvector; operations which require negligible time. The same remarks apply to the estimation of Lyapunov exponents in the following sections, where the variation of the exponents with respect to changing probabilities may be efficiently investigated with our method.

4 Lyapunov exponents

Once we have a rigorous estimate of the invariant measure of our random dynamical system, we may use it directly to compute dynamical indicators such as Lyapunov exponents. We assume that each of our maps T_k is a diffeomorphism so that we may assign k invertible Jacobian matrices $D_x T_k$, $k = 1, \ldots, r$ to each point $x \in M$. Throughout the remainder of the paper, we assume M is a compact subset of $\mathbb{R}^d$; in such a setting, $\mathcal{M}(M)$ is also compact and the metric d_H generates weak convergence on $\mathcal{M}(M)$. In this introductory section, we review the theory of Lyapunov exponents for a general random dynamical system formed by an iid composition of differentiable mappings.

Define $D_{(\omega,x)}T = D_x T_{\omega_0}$, and put

$$D_{(\omega,x)}T^N = D_{\tau_1^{N-1}(\omega,x)}T \circ \cdots \circ D_{\tau_1(\omega,x)}T \circ D_{(\omega,x)}T.$$

As we iterate random orbits of our system we repeatedly multiply together the Jacobian matrices and consider the limit:

$$\lim_{N\to\infty} \frac{1}{N} \log \left\| D_{(\omega,x)}T^N(v) \right\|. \tag{18}$$

In the sequel, we shall extensively use real projective space, namely the factor space $\mathbb{R}^d/\sim$, where two vectors $u \sim v \iff u = a \cdot v$ for some scalar $a \in \mathbb{R}$. We denote $\mathbb{R}^d/\sim$ by $\mathbb{RP}^{d-1}$; invertible $d \times d$ matrices have a natural action on $\mathbb{RP}^{d-1}$ under matrix multiplication. The following result of Kifer guarantees that these limits exist and are independent of the random sequence $\omega \in \Omega$.

Theorem 4.1 (Thm V.1.1 [11]): *Let μ be the unique $\mathcal{P}$-invariant measure on M, and suppose that*

$$\int_M \log^+ \|D_x T_{\omega_0}\| \, d\mu(x) \; < \infty,$$

for every $\omega_0 \in S$. Then for $\mathbb{P} \times \mu$-almost all (ω, x) one has that

(i) the limit (18) exists and takes on finitely many (non-random) values $-\infty < \lambda^{(m-1)} < \cdots < \lambda^{(0)} < \infty$, as v is varied over $\mathbb{R}^d$.

(ii) there exists a pointwise filtration of (random) subspaces, $\{0\} \subset V^{(m-1)}_{(\omega,x)} \subset \cdots \subset V^{(0)}_{(\omega,x)} = \mathbb{R}^d$ $(1 \le m \le d)$ such that the limit in (18) equals $\lambda^{(i)}$ if $v \in V^{(i)}_{(\omega,x)} \setminus V^{(i+1)}_{(\omega,x)}$

(iii) The subspaces $V^{(i)}_{(\omega,x)}$ satisfy

$$D_x T_{\omega_0}(V^{(i)}_{(\omega,x)}) = V^{(i)}_{\tau_1(\omega,x)} \tag{19}$$

(iv) Denote by Θ_i, the (non-random) value $\dim V^{(i)}_{(\omega,x)} - \dim V^{(i+1)}_{(\omega,x)}$ (so that $\sum_{i=0}^{m-1} \Theta_i = d$). Θ_i is the multiplicity of the exponent $\lambda^{(i)}$. If $D^{\wedge p}_{(\omega,x)}T$ denotes the p^{th} exterior power of $D_{(\omega,x)}T$, $(1 \le p \le d)$ then

$$\lim_{N\to\infty} \frac{1}{N} \log \|D^{\wedge p}_{(\omega,x)} T^N\| \tag{20}$$

exists and equals the sum of the p largest values $\lambda^{(i)}$, counted with multiplicity.

We will concern ourselves with estimating the top exponent $\lambda^{(0)}$, and recovering the remaining exponents by working with exterior powers.

4.1 Lyapunov exponents of compositions of affine maps/random matrix products

We begin by treating compositions of affine mappings (those with constant derivative) as they are more elementary to deal with, and have direct application to calculating Lyapunov exponents for iid random matrix products.

4.1.1 Space averages: Theory

Since the Jacobian matrices $D_x T_k$ are independent of x, we drop the subscript x in this section, and simply write DT_k. Clearly, the value of the limit in (18) is also independent of x, so the Lyapunov exponents depend only on the sequence ω, and the orbit of the random system in M is of no importance in the calculation. We shall use the following result of Furstenberg and Kifer [7] (modified for our purposes).

Theorem 4.2 ([7]): *Suppose that both $\|DT_{\omega_0}\|$ and $\|DT^{-1}_{\omega_0}\|$ are finite for each $\omega_0 \in S$, and that the only subspace of $\mathbb{R}^d$ that is invariant under all of the DT_{ω_0} is the trivial subspace. Then with probability one,*

$$\lim_{N\to\infty} \frac{1}{N} \log \|DT_{\omega_{N-1}} \circ \cdots \circ DT_{\omega_0}(v)\| = \lambda^{(0)} \tag{21}$$

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

$$\lambda^{(0)} = \lambda(\xi) := \sum_{k=1}^{r} w_k \int_{\mathbb{R}\mathrm{P}^{d-1}} \log \|DT_k(v)\| \, d\xi(v) \tag{22}$$

where $\xi \in \mathcal{M}(\mathbb{R}\mathrm{P}^{d-1})$ is any probability measure satisfying $\xi = \sum_{k=1}^{r} w_k \xi \circ DT_k^{-1}$.

Proof: This follows from the discussion in [7] following the proof of Lemma 3.3. □

We point out that it is very rare for all of the DT_k to share a common nontrivial invariant subspace.

Lemma 4.3: *Put the set of $d \times d$ real-valued matrices in 1-1 correspondence with $\mathbb{R}^{d^2}$ via $A = [a_{ij}] \mapsto (a_{11}, \ldots, a_{1d}, a_{21}, \ldots, a_{2d}, \ldots\ldots, a_{d1}, \ldots, a_{dd})$. Let $I \subset \mathbb{R}^{d^2}$ denote the subset of $\mathbb{R}^{d^2}$ corresponding to invertible matrices. If the Jacobian matrices $DT_1, \ldots, DT_r$ are selected at random (with respect to d^2-dimensional Lebesgue measure) from some subset of I with positive, finite d^2-dimensional Lebesgue measure, then with probability one, the matrices share no non-trivial invariant subspace.*

PROOF: Represent a given $d \times d$ matrix DT by considering its real-valued entries as a point in $\mathbb{R}^{d^2}$. In this coordinate system, the determinant of a DT is a polynomial of order d in the d^2 variables, and hence the condition $\det DT = 0$ is a $d^2 - 1$-dimensional manifold in $\mathbb{R}^{d^2}$. Denote the subset of $\mathbb{R}^{d^2}$ representing invertible matrices by I; clearly I has full d^2-dimensional Lebesgue measure since it is the complement of a $d^2 - 1$-dimensional manifold.

Denote by V_i, an i-dimensional invariant subspace of DT. We may send V_i to the subspace $E_i = \{e_1, \ldots, e_i, 0, \ldots, 0\}$ via a change of basis transformation. Since DT preserved the space V_i, the representation of DT in this new basis (denoted $DT_{[E]}$) must preserve E_i. Simple considerations imply that in this new basis, $i(d-i)$ elements of $DT_{[E]}$ must be specified as 0. Thus in $\mathbb{R}^{d^2}$, the set corresponding to matrices preserving V_i is an $d^2 - i(d-i)$-dimensional hyperplane.

Let $X_V = \{x \in \mathbb{R}^{d^2} : x \text{ corresponds to a matrix preserving the subspace } V\}$. Each X_V has d^2-dimensional Lebesgue measure 0. Let $U = \bigcup\{X_V : V \text{ is a subspace preserved by } DT\}$; this is a *finite* union of Lebesgue measure zero sets, since a $d \times d$ matrix has at most $\binom{d}{i}$ invariant i-dimensional subspaces.

Thus, for another matrix to share an invariant subspace with a given Jacobian matrix DT, it must belong to U, a Lebesgue measure zero set. So if our matrices $DT_1, \ldots, DT_r$ are chosen at random, from some subset of finite, positive measure of I, there is zero probability of them sharing a non-trivial invariant subspace. □

As we are really just multiplying together the r matrices $DT_1, \ldots, DT_r$ in an iid fashion, the problem of calculating the Lyapunov exponents of this random composition is almost mathematically equivalent to the calculation of the invariant measure μ. We replace M with $\mathbb{RP}^{d-1}$, T_k with DT_k, and define the skew-product $\tau_2 : \Omega \times \mathbb{RP}^{d-1} \circlearrowleft$ by $\tau_2(\omega, v) = (\sigma\omega, DT_{\omega_0}(v))$, where $v \in \mathbb{RP}^{d-1}$. The skew-product τ_2 provides a deterministic description of the dynamics on $\mathbb{RP}^{d-1}$, and using Lemma A.3, we produce an operator $\mathcal{D} : \mathcal{M}(\mathbb{RP}^{d-1}) \circlearrowleft$ defined by,

$$\mathcal{D}\zeta = \sum_{k=1}^{r} w_k \zeta \circ DT_k^{-1}, \qquad \zeta \in \mathcal{M}(\mathbb{RP}^{d-1}). \tag{23}$$

in analogy with $\mathcal{P}$. In the following sections, we describe a method of rigorously approximating a fixed point of $\mathcal{D}$ and then apply (22) to calculate $\lambda^{(0)}$.

4.1.2 Space averages: Our approximation

We begin with a simple lemma that will be used to provide a mathematical framework for numerical estimates of $\mathcal{D}$-invariant probability measures.

Lemma 4.4: *Let (X, ϱ) be a compact metric space. Let $T : X \circlearrowleft$ be continuous, and $T_n : X \circlearrowleft$, $n = 1, 2, \ldots$ be a family of maps such that*

$$\varrho(T_n x, Tx) \to 0 \text{ uniformly as } n \to \infty. \tag{24}$$

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 (24), and the third term goes to zero by hypothesis. □

To apply this Lemma we will 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 4.4 be the operator $\mathcal{D} : \mathcal{M}(\mathbb{RP}^{d-1}) \circlearrowleft$ defined in (23), and set about (i) showing continuity of $\mathcal{D}$, (ii) finding approximations $\mathcal{D}_n$ of $\mathcal{D}$, and (iii) finding fixed points of the $\mathcal{D}_n$.

Continuity of $\mathcal{D}$

Lemma 4.5: *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 particular, the operator $\mathcal{D} : \mathcal{M}(\mathbb{RP}^{d-1}) \circlearrowleft$ is continuous with respect to the metric d_H.

PROOF: The proof runs along the lines of the proof of Lemma 3.4. □

Approximation of $\mathcal{D}$

Lemma 4.6: *Partition $\mathbb{RP}^{d-1}$ into a finite number of connected, measurable subsets $\{B_1, \ldots, B_m\}$, choose and fix a single point $v_i \in B_i$, $i = 1, \ldots, m$, and construct the r $m \times m$ transition matrices:*

$$D_{m,ij}(k) = \begin{cases} 1, & \text{if } DT_k(v_i) \in B_j, \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

Define an operator $\mathcal{D}_m : \mathcal{M}(\mathbb{RP}^{d-1}) \circlearrowleft$ by

$$\mathcal{D}_m \zeta = \sum_{j=1}^{m} \left(\sum_{i=1}^{m} \zeta(B_i) \sum_{k=1}^{r} w_k D_{m,ij}(k) \right) \delta_{v_j}. \tag{26}$$

One has that

$$d_H(\mathcal{D}_m\zeta, \mathcal{D}\zeta) \leq \left(1 + \sum_{k=1}^{r} w_k \ell_k \right) \max_{1 \leq j \leq m} \operatorname{diam}(B_j)$$

for all $\zeta \in \mathcal{M}(\mathbb{RP}^{d-1})$.

PROOF: This follows as in the proof of Lemma 3.11. □

Fixed points of $\mathcal{D}_m$

Corollary 4.7: *Suppose we construct a sequence of operators $\mathcal{D}_m$ from a sequence of partitions of $\mathbb{RP}^{d-1}$ with maximal elements of decreasing diameter. Each of the $\mathcal{D}_m$ has at least one fixed point $\xi_m \in \mathcal{M}(\mathbb{RP}^{d-1})$, namely,*

$$\xi_m = \sum_{j=1}^{m} d_{m,i} \delta_{v_i}, \tag{27}$$

where d_m is a $1 \times m$ fixed left eigenvector of $D_m = \sum_{k=1}^{r} w_k D_m(k)$. If $\tilde{\xi}$ is any limit of the sequence $\{\xi_m\}$, then $\tilde{\xi}$ is a fixed point of $\mathcal{D}$.

PROOF: The existence of a fixed left eigenvector d_m follows from the fact that D_m is a stochastic matrix for each m. The invariance of ξ_m under $\mathcal{D}_m$ follows as in Lemma 3.9, and the fact that $\tilde{\xi}$ is invariant under $\mathcal{D}$ follows from Lemmas 4.6, 4.5 and 4.4. □

4.1.3 Convergence results

The following general lemma will simply say that we needn't worry about finding convergent subsequences for our estimates of $\lambda^{(0)}$.

Lemma 4.8: *Let $x_1, x_2, \ldots$ be an infinite sequence of points in a compact metric space (X, ϱ), and set $L \subset X$ to be the set of limits attained from all convergent subsequences of the fixed sequence $x_1, x_2, \ldots$. Suppose that $f : X \to \mathbb{R}$ is a continuous function such that $f(x) = c$ for all $x \in L$ and some constant $c \in \mathbb{R}$. Then $\lim_{n\to\infty} f(x_n)$ exists and equals c.*

PROOF: Straightforward. □

We may now use Corollary 4.7 and Theorem 4.2 to immediately obtain a method of approximation for the top Lyapunov exponent of our affine random dynamical system.

Proposition 4.9 (Main Result 2): *Under the hypotheses of Theorem 4.2, suppose that $\{\xi_m\}$ are a sequence of fixed points of the operators $\mathcal{D}_m$ as in (27). Then*

$$\sum_{k=1}^{r} w_k \sum_{i=1}^{m} d_{m,i} \log \|DT_k(v_i)\| := \lambda_m \to \lambda^{(0)} \qquad \text{as } m \to \infty. \tag{28}$$

PROOF:

We begin by noting that

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

(ii) the map $\lambda : \mathcal{M}(\mathbb{RP}^{d-1}) \to \mathbb{R}$ defined by $\lambda(\zeta) = \sum_{k=1}^{r} w_k \int_{\mathbb{RP}^{d-1}} \log \|DT_k(v)\| \, d\zeta(v)$ is continuous,

(iii) $\lambda(\xi) = \lambda^{(0)}$ for every $\xi \in \mathcal{M}(\mathbb{RP}^{d-1})$ satisfying $\mathcal{D}\xi = \xi$, (by Theorem 4.2),

(iv) $\lambda(\tilde{\xi}) = \lambda^{(0)}$, where $\tilde{\xi}$ is the limit of any weakly convergent subsequence of $\xi_1, \xi_2, \ldots$ by (iii) and Corollary 4.7,

Denoting the collection of all $\tilde{\xi}$ by Ξ, we apply Lemma 4.8 with $(X, \varrho) = (\mathcal{M}(\mathbb{RP}^{d-1}), d_H)$, $x_m = \xi_m$, $L = \Xi$, $f(\cdot) = \lambda(\cdot)$, and $c = \lambda^{(0)}$ to obtain the desired result. □

4.1.4 Numerical examples: Space average vs iterative methods

As we have no analytic error bounds for our approximate Lyapunov exponents, if we wish to make a comparison between our method and the random iteration approach, we must restrict ourselves to systems for which the Lyapunov exponent is known exactly. The class of such systems is very small indeed, and this is the main reason why it is important to develop rigorous and efficient approximation techniques such as the one introduced in this paper.

Example 4.10: As a first example, we choose the following system:

$$A_0 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Each of the above matrices is applied with equal probability 1/3; that is, $S = \{0, 1, 2\}$ and $\rho(\{k\}) = 1/3$, $k = 0, 1, 2$. The top Lyapunov exponent may be simply calculated analytically as $\lambda^{(0)} = (\log 2)/3$. This system has application in calculating the dimensions of intersections of Cantor sets [10]. The invariant probability measure $\xi \in \mathcal{M}(\mathbb{RP}^1)$ for this random matrix product is $(\delta_0 + \delta_{\pi/2})/2$; that is, half of the weight is given to the vector $\begin{pmatrix} 1 & 0 \end{pmatrix}$ and half is given to the vector $\begin{pmatrix} 0 & 1 \end{pmatrix}$.

Random iteration By putting $v = (\,1 \quad 0\,)$ in (21) we may rewrite this equation as

$$\begin{aligned}\lambda^{(0)} &= \lim_{N\to\infty} \frac{1}{N} \log \|A_{\omega_{N-1}} \circ \cdots \circ A_{\omega_0} (\,0 \quad 1\,)^\dagger\,\| \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \log \|A_{\omega_i} \circ \overline{A_{\omega_{i-1}} \circ \cdots \circ A_{\omega_0} (\,0 \quad 1\,)^\dagger}\,\| \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} H_i,\end{aligned}$$

where $\overline{v}$ denotes the normalisation of v, and H_i is a Bernoulli random variable, taking on the values $\log 2$ and 0, with probability 1/3 and 2/3, respectively. Our ability to write $\lambda^{(0)}$ as the limit of partial sums of Bernoulli random variables follows easily from the action of the matrices on the vectors $(\,1 \quad 0\,)$ and $(\,0 \quad 1\,)$. We may now subject the partial sums $S_N = \frac{1}{N}\sum_{i=0}^{N-1} H_i$ to the same analysis as in §3.3.1 to obtain similar bounds for the expected difference of S_N from $\lambda^{(0)} = \log(2)/3$. That is, $\mathbb{E}(|S_N - (\log 2)/3|) = \sum_{i=0}^{N} \left|\frac{i}{N}\log 2 - \frac{\log 2}{3}\right| \cdot \binom{N}{i}\left(\frac{1}{3}\right)^i \left(\frac{2}{3}\right)^{N-i}$. For $N = 60, 150, 300, 600$, these values are shown in Table 1. For $N = 1500, 3000$, we have approximate results based on Central Limit Theorem-type estimates.

Table 1: Lyapunov exponent estimates

# of partition sets m	# of iterations used $N = 3m$	Expected random iteration error $\mathbb{E}(\|S_N - (\log 2)/3\|)$	Estimate Error $\|\lambda_m - (\log 2)/3\|$
20	60	3.3×10^{-2}	2.3×10^{-3}
50	150	2.1×10^{-2}	3.7×10^{-4}
100	300	1.5×10^{-2}	9.2×10^{-5}
200	600	1.1×10^{-2}	2.3×10^{-5}
500	1500	$\approx 6.7 \times 10^{-3}$	3.7×10^{-6}
1000	3000	$\approx 4.7 \times 10^{-3}$	9.3×10^{-7}

Our method For $m = 20, 50, 100, 200, 500, 1000$, we construct three matrices $D_m(k)$, $k = 0, 1, 2$, where $D_m(k)$ is an approximation of the action of A_k on $\mathbb{RP}^1$. Using (28) we produce estimates λ_m of $\lambda^{(0)} = \log(2)/3$, shown in Table 1. These estimates are compared with plots of the partial sums S_N on an iteration for iteration basis ($N = 3m$) in Figure 3. From log-log fits of columns 3 and 4 of Table 1, the error in our estimates appears to be $O(m^2)$, while we expect the errors from the random iteration method to be $O(m^{1/2})$.

This first example is not really a fair comparison, as the invariant measure ξ is easily approximated by the eigenvectors d_m because of its simple structure.

Example 4.11 (Schrödinger's Equation and the Lloyd Model): We now apply our technique to a more complicated system; this second example also illustrates how one may deal with a product of matrices being drawn from an uncountably infinite collection.

This particular matrix product arises from a discrete one-dimensional Schrödinger equation on an infinite lattice $\{\ldots, -n, \ldots, -1, 0, 1, \ldots, n, \ldots\}$. The model is defined by the difference equation

$$\psi_{n+1} + \psi_{n-1} = (E - q_n)\psi_n,$$

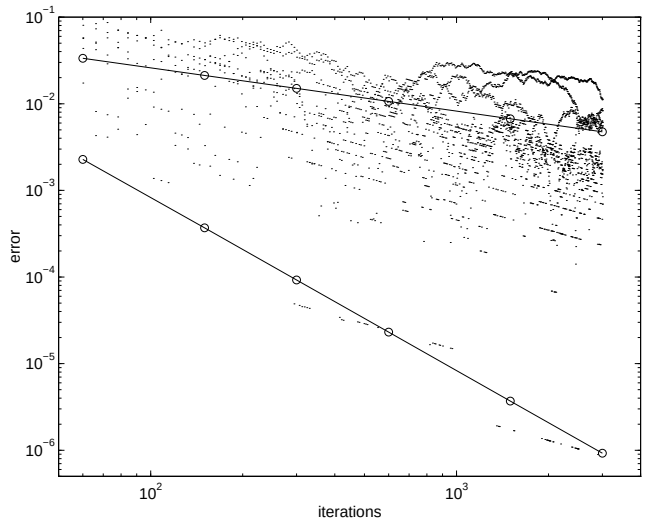

Figure 3: (i) Time average results of the error $|S_N - \log(2)/3|$ for 7 random orbits of length 3000 vs. N (dotted), (ii) expected value of the error $|S_N - \log(2)/3|$ as averaged over all possible orbits vs. N, $N = 60, \ldots, 3000$ (upper line, circles), and (iii) the errors of our estimates: $|\lambda_m - \log(2)/3|$ vs. $3m$, $m = 20, \ldots, 1000$ (lower line, circles)

where E is the "energy", and ψ_n and q_n are the amplitude of the wavefunction and the random potential, respectively, on lattice site n; see [3] §5.1 for details. The random potentials q_n are iid random variables. By rewriting this recursion formula in matrix form

$$\begin{pmatrix} \psi_{n+1} \\ \psi_n \end{pmatrix} = \begin{pmatrix} E - q_n & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \psi_n \\ \psi_{n-1} \end{pmatrix}, \tag{29}$$

we may study the growth of recursive solutions through a random matrix product. If such recursive solutions are valid, the *localisation length*,

$$L = -\lim_{|n|\to\infty} \frac{1}{|n|} < \log|\psi_n| >$$

is a measure of how the relative amplitudes of the elements of the eigenstate $\Psi = \{\ldots, \psi_{-n}, \ldots, \psi_0, \ldots, \psi_n, \ldots\}$ vary with n. The quantity L is simply the inverse of the top Lyapunov exponent of an iid random product of the matrices on the RHS of (29).

We choose $E = 2$, and at each time step select the q_n from a probability distribution with density $\phi(q) = 1/\pi(1+q^2)$. That is, low values of q are more likely to be selected than high positive or negative values. Such a choice of density gives rise to the "Lloyd model", and the top Lyapunov exponent may be expressed analytically as $\lambda^{(0)} = \cosh^{-1}\left((\sqrt{17}+1)/4\right)$; see [8], p.97.

Remark 4.12: Much is known about the Lyapunov exponents of the random matrix product (29). Our primary reason for selecting the Lloyd model is the existence of an analytic value for $\lambda^{(0)}$. This allows an exact numerical evaluation of the error in our method and a comparison with the standard iterative method. It is not suggested that either a numerical space average or time average be used to calculate exponents for this well-studied system.

Operationally, to generate a random value of q according to the distribution ϕ, we integrate ϕ to produce $\Phi(q) = (\tan^{-1} q)/\pi + 1/2$, and invert to obtain $\Phi^{-1}(q) = \tan(\pi(q - 1/2))$.

The mapping $\Phi^{-1} : [0,1] \to \mathbb{R}$ takes the uniform density on $[0,1]$ to the density of ϕ on $\mathbb{R}$. In order to do calculations on a computer, we need to approximate our infinite set of matrices with a finite set. We choose a simple minded method, namely dividing $\mathbb{R}$ into r intervals of equal probability: $\Phi^{-1}([(k-1)/r, k/r))$, $k = 1, \ldots, r$ and selecting values of q from the images of the midpoints: $q_k = \Phi^{-1}((2k-1)/2r)$, $k = 1, \ldots, r$. In this way, we produce an approximate iid random composition of a *finite* set of matrices $\{M_{q_k} : k = 1, \ldots, r\}$, with corresponding probabilities $w_k = 1/r$. The action of each of these matrices on $\mathbb{RP}^{d-1}$ will be approximated as before, by partitioning $\mathbb{RP}^{d-1}$ into m sets. Thus there are two variables in our approximation; the number of matrices used r, and the number of partition sets m. We find that r plays a larger role in the calculation of the Lyapunov exponents. The reason for this is that very negative and very positive values of q, while unlikely, produce large contributions to the exponent calculation. Thus, for accurate approximations, it is important to include many of these outlying matrices, and perhaps there are better ways of forming the finite set $\{M_{q_k} : k = 1, \ldots, r\}$ than our simple minded method.

Using 1600 matrices selected as described above, and partitioning $\mathbb{RP}^{d-1}$ into 100 equal subintervals, we produce 1600 100×100 matrices $D_{100}(k)$, $k = 1, \ldots, 1600$, as in (25), and combine them to form $D_{100} = \sum_{k=1}^{1600} D_{100}(k)/1600$. A left eigenvector $d_{100}^{(1600)}$ of D_{100} is calculated, and a probability measure $\xi_{100}^{(1600)} = \sum_{i=1}^{100} d_{100,i}^{(1600)}$ produced as in (27). The measure $\xi_{100}^{(1600)}$ is an approximate fixed point of $\mathcal{D}$; see Figure 4. To produce this approximate fixed

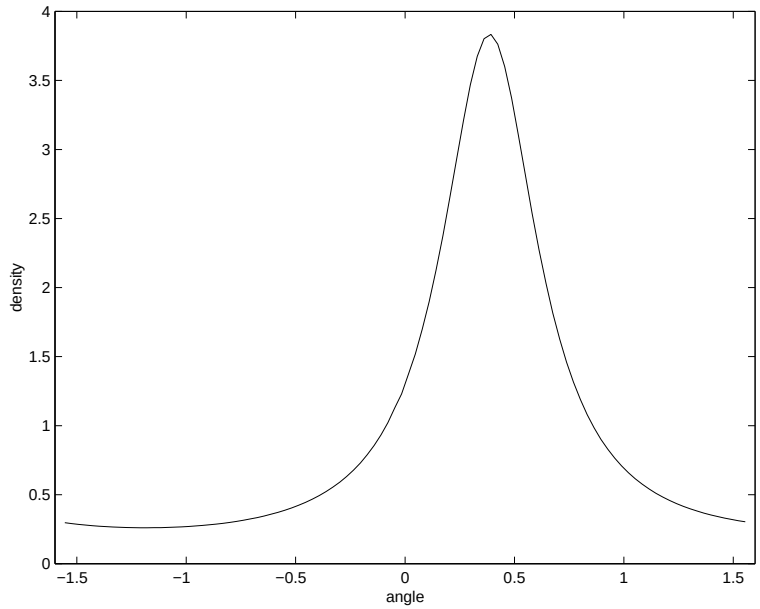

Figure 4: Density of the approximate $\mathcal{D}$-invariant measure $\xi_{100}^{(1600)}$. This plot is a linear interpolation of a step function that takes the value $100 d_{100,i}^{(1600)}$ on the set $[-\pi/2 + (i-1)\pi/100, -\pi/2 + i\pi/100)$, $i = 1, \ldots, 100$.

point, we used 100×1600 iterations. The Lyapunov exponent estimates for differing values of r are shown in Table 2. Our estimates monotonically increase with r because of the increasing use of matrices M_q with outlying values of q. The error in our estimates appears to decrease like $O(1/r)$.

Figure 5 compares our estimates with those from 10 random orbits using (21), on an iteration for iteration basis. Our construction used $100r$ iterations to produce the estimate $\lambda_{100}^{(r)}$, so we compare this value with random iteration estimates after $100r$ iterations. Even

Table 2: Lyapunov exponents estimates using a space average on a 100 set partition of $\mathbb{R}\mathrm{P}^{d-1}$.

Number of Matrices r	Number of Iterations $100r$	Estimate $\lambda_{100}^{(r)}$	Error $\lvert\lambda^{(0)} - \lambda_{100}^{(r)}\rvert$
50	5000	0.7192	1.4×10^{-2}
100	10000	0.7256	7.1×10^{-3}
200	20000	0.7292	3.5×10^{-3}
400	40000	0.7311	1.8×10^{-3}
800	80000	0.7320	9.0×10^{-4}
1600	160000	0.7324	4.6×10^{-4}
3200	320000	0.7326	2.6×10^{-4}

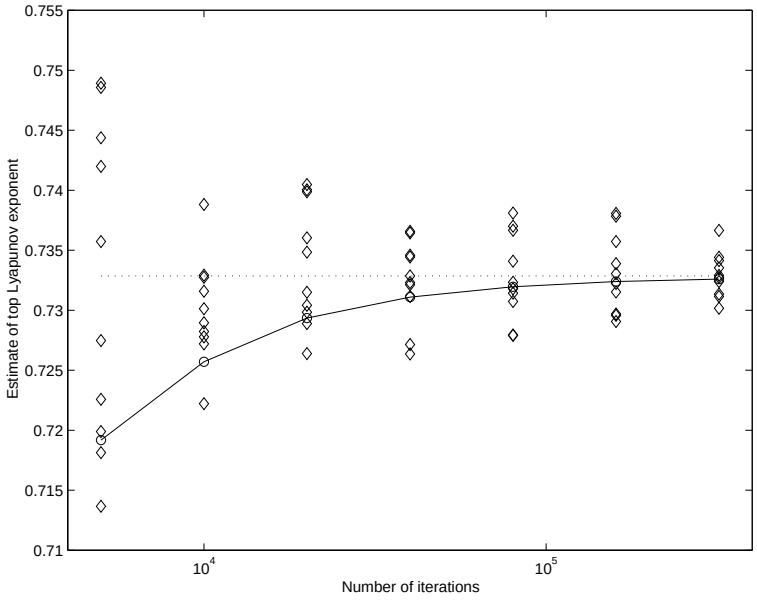

Figure 5: Plot of Lyapunov exponent estimates from 10 random orbits vs. number of iterations, $N = 5000, 10000, 20000, 40000, 80000, 160000, 320000$ (diamonds), and Lyapunov exponent estimates $\lambda_{100,r}$ from the space average vs. $N = 100r$ (solid line and circles). The exact answer is shown dotted.

after 320000 iterations, there is still significant fluctuation in the time average results.

Further numerical results are detailed in [6].

4.2 Lyapunov exponents of nonaffine IFS's

4.2.1 Space averages: Theory

As our Jacobian matrices now depend on $x \in M$, we adjoin M to real projective space $\mathbb{R}\mathrm{P}^{d-1}$ and define the mappings $\mathcal{T}_k : M \times \mathbb{R}\mathrm{P}^{d-1} \circlearrowright$, $k = 1, \ldots, r$, by $\mathcal{T}_k(x, v) = (T_k x, D_x T_k(v))$; clearly $\mathcal{T}_k^{-1}$ is given by $\mathcal{T}_k^{-1}(x, v) = (T_k^{-1} x, (D_{T_k^{-1}x} T)^{-1}(v))$. We now form the skew product $\tau_3 : \Omega \times (M \times \mathbb{R}\mathrm{P}^{d-1}) \circlearrowright$, by $\tau_3(\omega, (x, v)) = (\sigma\omega, \mathcal{T}_{\omega_0}(x, v))$. We again wish to find an invariant probability measure for τ_3, and by Lemma A.3 are able to construct one using fixed points

of the operator $\mathcal{D}' : \mathcal{M}(M \times \mathbb{R}\mathrm{P}^{d-1}) \circlearrowleft$ defined by

$$\mathcal{D}'\zeta' = \sum_{k=1}^{r} w_k \zeta' \circ \mathcal{T}_k^{-1}, \quad \zeta' \in \mathcal{M}(M \times \mathbb{R}\mathrm{P}^{d-1}). \tag{30}$$

Let $\pi : M \times \mathbb{R}\mathrm{P}^{d-1} \to M$ denote the projection onto the first coordinate of $M \times \mathbb{R}\mathrm{P}^{d-1}$. It is clear that if a measure $\xi' \in \mathcal{M}(M \times \mathbb{R}\mathrm{P}^{d-1})$ is $\mathcal{D}'$-invariant, then the measure $\xi' \circ \pi^{-1}$ on M is $\mathcal{P}$-invariant, and by uniqueness, $\xi' \circ \pi^{-1} = \mu$. Since M and $\mathbb{R}\mathrm{P}^{d-1}$ are compact metric spaces, we may uniquely (for μ-almost all x) disintegrate ξ' as

$$\xi'(E) = \int_M \xi_x(E_x)\, d\mu(x), \tag{31}$$

where the section $E_x = \{v \in \mathbb{R}\mathrm{P}^{d-1} : (x, v) \in E\}$.

The appropriate version of Theorem 4.2 for nonaffine IFS is as follows:

Theorem 4.13 ([11]): *Let μ be the unique $\mathcal{P}$-invariant probability measure and assume that*

$$\int_M \log^+ \|D_x T_{\omega_0}\|\, d\mu(x) < \infty$$

for each $\omega_0 \in S$. In addition, let $\Omega' \times M' \subset \Omega \times M$ be the subset of full $\mathbb{P} \times \mu$ measure for which the conclusions of Theorem 4.1 hold, and suppose that

$$\bigcap_{\omega \in \Omega'} V^{(1)}_{(\omega,x)} = \{0\}. \tag{32}$$

Then for $\mathbb{P} \times \mu$-almost all (ω, x),

$$\lim_{N\to\infty} \frac{1}{N} \log \|D_{(\omega,x)} T^N(v)\| = \lambda^{(0)}$$

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

$$\lambda^{(0)} = \lambda(\xi') := \sum_{k=1}^{r} w_k \int_M \int_{\mathbb{R}\mathrm{P}^{d-1}} \log \|D_x T_k(v)\|\, d\xi_x(v)\, d\mu(x) \tag{33}$$

where $\xi' = \int_M \xi_x\, d\mu(x)$ as in (31) and $\xi' \in \mathcal{M}(M \times \mathbb{R}\mathrm{P}^{d-1})$ is any $\mathcal{D}'$-invariant probability measure.

Proof: Follows from Theorems III.1.2 and III.1.3 [11]. □

Remark 4.14: The condition that the largest proper subspaces $V^{(1)}_{(\omega,x)}$ have trivial intersection produces the same effect as the condition in the affine case (Theorem 4.2) that the Jacobian matrices share no nontrivial invariant subspace. As in the affine case, it is usual for this condition to be satisfied, and hence that the random system has only one observable Lyapunov exponent.

4.2.2 Space averages: Our approximation

Denote by κ the natural Riemannian metric on $\mathbb{RP}^{d-1}$. We begin by noting that since each T_k is at least C^1 and M is compact, the Jacobian matrices have the property that there exist constants $\operatorname{Lip}_x DT_k < \infty$ for each $k = 1, \ldots, r$, such that

$$\kappa(D_{x_1} T_k(v), D_{x_2} T_k(v)) \leq \operatorname{Lip}_x DT_k \cdot d(x_1, x_2) \qquad \text{for all } x \in M \text{ and } k = 1, \ldots, r,$$

Lemma A.3 gives us a simple way to represent τ_3-invariant measures. We will use Lemma 4.4 to approximate fixed points of (30). Firstly, we must show that the operator $\mathcal{D}'$ is continuous, and secondly that our approximation $\mathcal{D}'_{m,n}$ satisfies $\mathcal{D}'_{m,n}\zeta' \to \mathcal{D}'\zeta'$ for ζ' in a suitable subset of $\mathcal{M}(M \times \mathbb{RP}^{d-1})$.

Continuity of $\mathcal{D}'$

Lemma 4.15: *Let $\operatorname{Lip} T_k$ denote the Lipschitz constant of T_k (using the metric d), and $\operatorname{Lip} D_x T_k$ denote the Lipschitz constant of $D_x T_k$ (using the metric κ). Define a metric $\tilde{\kappa}$ on $M \times \mathbb{RP}^{d-1}$ by $\tilde{\kappa}((x_1, v_1), (x_2, v_2)) = \max\{d(x_1, x_2), \kappa(v_1, v_2)\}$, and denote by $\operatorname{Lip}(s)$ the space of real-valued functions on $M \times \mathbb{RP}^{d-1}$ with Lipschitz constant s (using the metric $\tilde{\kappa}$). Then*

$$d_H(\mathcal{D}'\zeta'_1, \mathcal{D}'\zeta'_2) \leq \max_{1 \leq k \leq r} \left(\max\{\operatorname{Lip} T_k, \sup_{x \in M} \operatorname{Lip} D_x T_k\} \right) d_H(\zeta'_1, \zeta'_2) \tag{34}$$

for any $\zeta'_1, \zeta'_2 \in \mathcal{M}(M \times \mathbb{RP}^{d-1})$

Proof: As in the proof of Lemma 4.5. □

Approximation of $\mathcal{D}'$

Definition 4.16: We partition M into n (connected, measurable) sets $A_1, \ldots, A_n$ and $\mathbb{RP}^{d-1}$ into m (connected, measurable) sets $B_1, \ldots, B_m$.

From each set A_i, choose a single point a_i, $i = 1, \ldots, m$, and define an $n \times n$ stochastic matrix $P_n(k)$ by setting

$$P(k)_{n,ij} = \begin{cases} 1, & \text{if } T_k a_i \in A_j, \\ 0, & \text{otherwise} \end{cases}. \tag{35}$$

From each set B_g, choose a single point b_g, $g = 1, \ldots, n$. For each mapping $D_{a_i} T_k : \mathbb{RP}^{d-1} \circlearrowleft$, $k = 1, \ldots, r$, $i = 1, \ldots, n$, define an $m \times m$ stochastic matrix $D_m(k, i)$ by setting

$$D_{m,gh}(k, i) = \begin{cases} 1, & \text{if } D_{a_i} T_k(b_g) \in B_h, \\ 0, & \text{otherwise} \end{cases}. \tag{36}$$

Further, define the class of measures $\mathfrak{M}_n \subset \mathcal{M}(M \times \mathbb{RP}^{d-1})$ by

$$\mathfrak{M}_n = \{\zeta' \in \mathcal{M}(M \times \mathbb{RP}^{d-1}) : \zeta' = \int_M \zeta_x \, d\nu(x), \text{ where } \nu \in \mathcal{M}(M) \text{ and } \zeta_x = \zeta_i \text{ for } x \in A_i\} \tag{37}$$

That is, we restrict our attention to measures in $\mathcal{M}(M \times \mathbb{RP}^{d-1})$ which may be disintegrated in such a way that the conditional measures on $\mathbb{RP}^{d-1}$ (conditioned on $x \in M$) do not vary over each partition set A_i. We may write an element of $\mathfrak{M}_n$ as

$$\zeta'(E) = \int_M \zeta_{i_x}(E_x) \, d\nu(x), \tag{38}$$

where $i_x = i$ if $x \in A_i$, and $E_x = \{v \in \mathbb{R}\mathrm{P}^{d-1} : (x,v) \in E\}$. We now define the operator $\mathcal{D}'_{n,m} : \mathfrak{M}_n \circlearrowleft$ which will approximate $\mathcal{D}'$ on $\mathfrak{M}_n$. Let ζ' have representation as in (38).

$$\mathcal{D}'_{n,m}\zeta' = \sum_{j=1}^{n}\sum_{h=1}^{m}\left(\sum_{k=1}^{r} w_k \left(\sum_{i=1}^{n}\nu(A_i)P_{n,ij}(k)\sum_{g=1}^{m}\zeta_i(B_g)D_{m,gh}(k,i)\right)\right)\delta_{(a_j,b_h)} \tag{39}$$

Proposition 4.17:

$$d_H(\mathcal{D}'_{n,m}\zeta', \mathcal{D}'\zeta') \to 0 \qquad \textit{uniformly as } n,m \to \infty.$$

Proof: As in the proof of Lemma 4.6. □

Fixed points of $\mathcal{D}'_{n,m}$ We now show that $\mathcal{D}'_{n,m}$ has a fixed point in $\mathfrak{M}_n$. From the form of (39), it is clear that a fixed point must be a convex combination of δ-measures at the points $(a_j, b_h),\ 1 \le j \le n,\ 1 \le h \le m$. We may therefore write such a measure as

$$\xi'_{n,m} = \sum_{i=1}^{n} t_{n,i} \sum_{g=1}^{m} s^{(i)}_{m,g}\delta_{(a_i,b_g)}, \tag{40}$$

for suitable nonnegative real numbers $t_{n,i},\ 1 \le i \le n$ and $s^{(i)}_{m,g},\ 1 \le i \le n,\ 1 \le g \le m$. By defining measures $\mu_n = \sum_{i=1}^{n} t_{n,i}\delta_{a_i}$ and $\xi_{n,m,i} = \sum_{g=1}^{m} s^{(i)}_{m,g}\delta_{b_g}$, on M and $\mathbb{R}\mathrm{P}^{d-1}$, respectively, we may represent $\xi'_{n,m}$ as

$$\xi'_{n,m}(E) = \sum_{i=1}^{n}\int_{A_i} \xi_{n,m,i}(E_x)\, d\mu_n(x) \in \mathfrak{M}_n.$$

We now begin to put constraints on the t's and s's. Firstly, we note that

$$\xi'_{n,m}(A_i \times \mathbb{R}\mathrm{P}^{d-1}) = \mu_n(A_i) = t_{n,i}, \tag{41}$$

so that

$$\sum_{i=1}^{n} t_{n,i} = 1. \tag{42}$$

Secondly, note that $\xi'_{n,m}(A_i \times B_g) = \mu_n(A_i)\xi_{n,m,i}(B_g) = t_{n,i}s^{(i)}_{m,g}$ so that

$$\xi_{n,m,i}(B_g) = s^{(i)}_{m,g}. \tag{43}$$

Thus

$$\sum_{g=1}^{m} s^{(i)}_{m,g} = 1 \text{ for all } i = 1,\ldots,n. \tag{44}$$

These two conditions are simply for normalisation purposes; to find suitable values for $t_{n,i}$ and $s^{(i)}_{m,g}$, we must appeal to the invariance property of $\xi'_{n,m}$.

We compute that

$$\mathcal{D}'_{n,m}\xi'_{n,m} = \sum_{k=1}^{r} w_k \sum_{j=1}^{n}\sum_{h=1}^{m}\left(\sum_{i=1}^{n} t_{n,i}P_{n,ij}(k)\sum_{g=1}^{m} s^{(i)}_{m,g}D_{m,gh}(k,i)\right)\delta_{(a_j,b_h)}$$

from (39), (41) and (43). Equating $\mathcal{D}'_{n,m}\xi'_{n,m}(A_j \times \mathbb{R}\mathrm{P}^{d-1})$ with $\xi'_{n,m}(A_j \times \mathbb{R}\mathrm{P}^{d-1})$ gives $\sum_{k=1}^{r} w_k \sum_{i=1}^{n} t_{n,i}P_{n,ij}(k) = t_{n,j}$. Thus t_n is a fixed left eigenvector of $\sum_{k=1}^{r} w_k P_n(k)$. Equating $\mathcal{D}'_{n,m}\xi'_{n,m}(A_j \times B_h)$ with $\xi'_{n,m}(A_j \times B_h)$ yields

$$\sum_{k=1}^{r} w_k \sum_{i=1}^{n} t_{n,i}P_{n,ij}(k) \sum_{g=1}^{m} s^{(i)}_{m,g} D_{m,gh}(k,i) = t_{n,j}s^{(j)}_{m,h}.$$

Rearranging, we have

$$s^{(j)}_{m,h} = \sum_{k=1}^{r} w_k \sum_{i=1}^{n} \underbrace{\left(\frac{t_{n,i}P_{n,ij}(k)}{t_{n,j}}\right)}_{P^*_{n,ji}(k)} \sum_{g=1}^{m} s^{(i)}_{m,g} D_{m,gh}(k,i), \tag{45}$$

where for brevity, we denote $P_{n,ij}(k)t_{n,i}/t_{n,j}$ by $P^*_n(k)$. Equation (45) may be written in matrix form as:

$$(\, s^{(1)}_m | s^{(2)}_m | \cdots | s^{(n)}_m \,) = (\, s^{(1)}_m | s^{(2)}_m | \cdots | s^{(n)}_m \,)\, \mathfrak{D}_{n,m}, \tag{46}$$

where

$$\mathfrak{D}_{n,m} = \sum_{k=1}^{r} w_k \left(\begin{array}{c|c|c|c} P^*_{n,11}(k)\ D_m(k,1) & P^*_{n,21}(k)\ D_m(k,1) & \cdots & P^*_{n,m1}(k)\ D_m(k,1) \\ \hline P^*_{n,12}(k)\ D_m(k,2) & P^*_{n,22}(k)\ D_m(k,2) & \cdots & P^*_{n,m2}(k)\ D_m(k,2) \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline P^*_{n,1m}(k)\ D_m(k,m) & P^*_{n,2m}(k)\ D_m(k,n) & \cdots & P^*_{n,mm}(k)\ D_m(k,n) \end{array}\right) \tag{47}$$

Note that each entry displayed in (47) is an $m \times m$ block. The required solution will satisfy (45) and the constraint (44). Using the formula $P^*_{n,ji}(k) = P_{n,ij}(k)t_{n,i}/t_{n,j}$, it is straightforward to show that the matrix in (46) is similar to a stochastic matrix and hence has a fixed left eigenvector. Note that in the case of affine systems, $P_{n,ij}(k) = p_{n,j}(k)$ (the fixed left eigenvector of $P_n(k)$) for all i, so $P^*_{n,ji}(k) = p_{n,i}(k)$ for all j, and the block rows in (46) are identical, so $s^{(i)}_n = s^{(j)}_n$ for all $1 \le i, j \le n$, and we recover the simpler affine formula.

The above considerations lead to

Lemma 4.18: *The operator $\mathcal{D}'_n : \mathfrak{M}_n \circlearrowleft$ has a fixed point, namely, the probability measure given by (40), where*

(i) t_n is a fixed left eigenvector of $\sum_{k=1}^{r} w_k P_n(k)$, and $\sum_{i=1}^{n} t_{n,i} = 1$, and

(ii) $(s^{(1)}_m | s^{(2)}_m | \cdots | s^{(n)}_m)$ is a fixed left eigenvector of $\mathfrak{D}_{n,m}$ and $\sum_{g=1}^{m} s^{(i)}_{m,g} = 1$ for all $i = 1, \ldots, n$.

4.2.3 Convergence results

Proposition 4.19 (Main Result 3): *Under the hypotheses of Theorem 4.13, let $t_{n,i}$ and $s^{(i)}_{m,g}$, $i = 1, \ldots, n$, $g = 1, \ldots, m$ be as described in Lemma 4.18. Then*

$$\sum_{k=1}^{r} w_k \sum_{i=1}^{n} t_{n,i} \sum_{g=1}^{m} s^{(i)}_{m,g} \log \|D_{a_i} T_k(b_g)\| := \lambda_{n,m} \to \lambda^{(0)} \qquad \text{as } n, m \to \infty. \tag{48}$$

Proof: Since

(i) $\mathcal{D}' : \mathcal{M}(M \times \mathbb{R}\mathrm{P}^{d-1}) \circlearrowleft$ is continuous with respect to the d_H metric (Lemma 4.15),

(ii) $d_H(\mathcal{D}'\zeta', \mathcal{D}'_{n,m}\zeta') \to 0$ uniformly as $n, m \to \infty$, $\zeta' \in \mathfrak{M}_n$ (Lemma 4.17),

(iii) $\mathcal{D}'_{n,m}$ has a fixed point $\xi'_{n,m} \in \mathfrak{M}_n$ for each $n, m \geq 0$ (Lemma 4.18),

we have by Lemma 4.4 that any limit $\tilde{\xi}'$ of a convergent subsequence of the sequence $\{\xi'_{n,m}\}$ is $\mathcal{D}'$-invariant.

Combining this result with the facts that

(i) the mapping $\lambda : \mathcal{M}(M \times \mathbb{R}\mathrm{P}^{d-1}) \to \mathbb{R}$ given by $\lambda(\zeta') = \sum_{k=1}^{r} w_k \int_{M \times \mathbb{R}\mathrm{P}^{d-1}} \log \|D_x T_k(v)\| \, d\zeta'(x, v)$ is continuous,

(ii) $\lambda(\xi') = \lambda^{(0)}$ for any $\mathcal{D}'$-invariant probability measure ξ' (by hypothesis),

we apply Lemma 4.8, with $(X, \varrho) = (\mathcal{M}(M \times \mathbb{R}\mathrm{P}^{d-1}), d_H)$, $f(\cdot) = \lambda(\cdot)$, $c = \lambda^{(0)}$, $x_n = \xi'_{n,m}$, and $L = \{\tilde{\xi}' \in \mathcal{M}(M \times \mathbb{R}\mathrm{P}^{d-1}) : \tilde{\xi}' \text{ is the limit of some convergent subsequence of } \{\xi'_{n,m}\}\}$, to obtain the desired result. □

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 $\mathfrak{D}_{n,m}$.

Corollary 4.20 (Main Result 3 in One Dimension): *Suppose that M is one-dimensional. Under the hypotheses of Theorem 4.13 (ignoring (32)), let $t_{n,i}$, $i = 1, \ldots, n$ be as described in Lemma 4.18. Then*

$$\sum_{k=1}^{r} w_k \sum_{i=1}^{n} t_{n,i} \log |D_{a_i} T_k| := \lambda_{n,m} \to \lambda^{(0)} \quad \text{as } n, m \to \infty. \tag{49}$$

4.3 Final Remarks

Under the conditions of Theorems 4.2 or 4.13 we may use our techniques to estimate the remaining (unobservable) Lyapunov exponents $\lambda^{(i)}$, $i = 1, \ldots, m-1$ in Theorem 4.1 by repeating the arguments in the proofs of Propositions 4.9 and 4.19 using exterior products of DT_k of ever increasing powers. These remaining exponents are observed with probability 0 when only the action of the Jacobian matrices on vectors (rather than higher dimensional subspaces) is considered.

5 Acknowledgements

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

A Discrete IID forcing and skew products

Here we detail the skew product representation of iid forcing of a finite collection of continuous maps. We will be using this representation in three instances to help us compute (i) the unique invariant probability measure μ of the random dynamical system, (ii) the Lyapunov exponents of an affine random dynamical system or random matrix product, and (iii) the Lyapunov exponents of a nonaffine random dynamical system. The main lemma of this section will simplify the representation of three objects:

(i) *(Invariant measure of the random system):* the unique probability measure $\mu \in \mathcal{M}(M)$ that is invariant on average under the action of our random system; in this case, the maps being forced are the $T_k : M \circlearrowleft$,

(ii) *(Lyapunov exponents of the affine random system or random matrix product):* a probability measure $\xi \in \mathcal{M}(\mathbb{RP}^{d-1})$ that is invariant on average under the action of the Jacobian matrices; in this case, the maps being forced are the constant Jacobian matrices $DT_k : \mathbb{RP}^{d-1} \circlearrowleft$.

(iii) *(Lyapunov exponents of the nonaffine random system):* a probability measure $\xi' \in \mathcal{M}(M \times \mathbb{RP}^{d-1})$ that is invariant on average under the action of the mappings $\mathcal{T}_k : M \times \mathbb{RP}^{d-1} \circlearrowleft$, introduced in §4.2.1; here it is the $\mathcal{T}_k$ that are being forced.

We show how one arrives at the intuitively plausible form of the Markov operator $\mathcal{P}$ from a general skew product representation of our random iid composition. The description here refers to the "Invariant measure of the random system" case, but with the appropriate substitutions also applies to the "Lyapunov exponents of the affine random system" and "Lyapunov exponents of the nonaffine random system" cases.

Definition A.1: A probability measure $\tilde{\mu}$ on $\Omega \times M$ is τ_1-*invariant* if:

(i) $\tilde{\mu} = \tilde{\mu} \circ \tau_1^{-1}$,

(ii) the projection of $\tilde{\mu}$ onto Ω is $\mathbb{P}$; that is, $\tilde{\mu}(F \times M) = \mathbb{P}(F)$ for all measurable $F \subset \Omega$.

When we talk about an invariant probability measure for our random system, what we really mean is an invariant measure "on average". The probability measure μ that is a fixed point of $\mathcal{P}$ may not be invariant under *any* of the transformations T_k, but the expectation (or average) of their combined actions *does* preserve μ. This motivates the following definition:

Definition A.2: A probability measure μ on M is called invariant under the iid composition $(\underset{\sim}{w}, \{T_k\})$ if $\mu(\cdot) = \tilde{\mu}(\Omega, \cdot)$, where $\tilde{\mu}$ is τ_1-invariant for the corresponding skew-product.

Thus we define an invariant measure for our IFS to be a measure that is invariant under the "expected" evolution of the system.

We now state a theorem due to Ohno [12] to make this precise.

Lemma A.3 (Ohno [12]): *A measure μ on M is $\mathcal{P}$-invariant iff the product measure $\mathbb{P} \times \mu$ is τ_1-invariant.*

Thus pushing forward $\tilde{\mu}$ under τ_1 and averaging with respect to $\mathbb{P}$ is the same as averaging with respect to $\mathbb{P}$ and pushing forward under $\mathcal{P}$; that is, $\tilde{\mu} \circ \tau_1^{-1}(\Omega, \cdot) = \mathcal{P}\mu(\cdot)$.

Lemma A.3 gives us a simple way to define τ_1-invariant measures using fixed points of the operator $\mathcal{P}$.

B Construction of the invariant probability measure, and a proof of $\mathcal{P}$ invariance

We construct a measure on Λ as follows. Firstly, define a family of disjoint closed subintervals $X_{i_0 \cdots i_{m-1}} \subset [0,1]$ ($i_k \in \{0,2\}$ for $0 \le k \le m-1$) as the set of points in the unit interval that have ternary expansions between (and including) $\cdot i_0 i_1 \cdots i_{m-1} 000 \cdots$ and $\cdot i_0 i_1 \cdots i_{m-1} 222 \cdots$. Secondly define a family of disjoint open subintervals $E_{i_0 \cdots i_{m-2} 1} \subset [0,1]$ ($i_k \in \{0,2\}$ for

$0 \leq k \leq m-2$) as $E_{i_0\cdots i_{m-2}1} = X_{i_0\cdots i_{m-2}} \setminus \left(X_{i_0\cdots i_{m-2}0} \cup X_{i_0\cdots i_{m-2}2}\right)$; these are the open subintervals that lie in between the closed subintervals $X_{i_0\cdots i_{m-1}}$. It may be shown that $\Lambda = \bigcap_{m=1}^{\infty} \left(\bigcup_{i_0\cdots i_{m-1}\in\{0,2\}} X_{i_0 i_1\cdots i_{m-1}}\right)$, where Λ is an intersection of a nested sequence of closed non-empty sets. We call a set $X_{i_0 i_1\cdots i_{m-1}}$ an *m-cylinder set*; the collection of all m-cylinder sets for finite m (we also include $\{\emptyset\}$ and Λ, the latter being a 0-cylinder set), together with the collection of sets $E_{i_0\cdots i_{m-2}1}$ for finite m, form a semi-algebra on Λ, denoted by $\mathcal{A}$. We define a function $\zeta : \mathcal{A} \to [0,1]$ by $\zeta(X_{i_0\cdots i_{m-1}}) = 1/2^m$, $\zeta(E_{i_0\cdots i_{m-2}1}) = 0$ and $\zeta(\{\emptyset\}) = 0$. Since ζ is countably additive, we may uniquely extend it to a probability measure on the σ-algebra generated by $\mathcal{A}$.

We now show that this probability measure μ is $\mathcal{P}$-invariant, where $\mathcal{P}\nu = (\nu \circ T_1^{-1} + \nu \circ T_2^{-1})/2$. By a standard result, it is enough to show invariance for sets in our semi-algebra $\mathcal{A}$. The important properties of our m-cylinders are:

(i) $T_1^{-1} X_{0 i_1\cdots i_{m-1}} = X_{i_1\cdots i_{m-1}}$,

(ii) $T_2^{-1} X_{0 i_1\cdots i_{m-1}} = \{\emptyset\}$,

(iii) $T_1^{-1} X_{2 i_1\cdots i_{m-1}} = \{\emptyset\}$,

(iv) $T_2^{-1} X_{2 i_1\cdots i_{m-1}} = X_{i_1\cdots i_{m-1}}$,

while the sets $E_{i_0\cdots i_{m-2}1}$ have inverses:

(i) $T_1^{-1} E_{0 i_1\cdots i_{m-2}1} = E_{i_1\cdots i_{m-2}1}$,

(ii) $T_2^{-1} E_{0 i_1\cdots i_{m-2}1} = \{\emptyset\}$,

(iii) $T_1^{-1} E_{2 i_1\cdots i_{m-2}1} = \{\emptyset\}$,

(iv) $T_2^{-1} E_{2 i_1\cdots i_{m-2}1} = E_{i_1\cdots i_{m-2}1}$,

We deal with the zero measure E-sets first. Clearly,

$$(\mathcal{P}\mu)(E_{i_0\cdots i_{m-2}1}) = \left(\mu(T_1^{-1} E_{i_0\cdots i_{m-2}1}) + \mu(T_2^{-1} E_{i_0\cdots i_{m-2}1})\right)/2 = 0,$$

as E-sets are transformed into either E-sets or the emptyset, both of which have measure zero. Since $\mu(E_{i_0\cdots i_{m-2}1}) = 0$, we have shown the invariance of μ on the E-sets. It is also clear that

$$(\mathcal{P}\mu)(X_{i_0\cdots i_{m-1}}) = (\mu(T_1^{-1} X_{i_0\cdots i_{m-1}}) + \mu(T_2^{-1} X_{i_0\cdots i_{m-1}}))/2$$

will produce a sum of $(1/2^{m-1} + 0)/2$ or $(0 + 1/2^{m-1})/2$, when i_0 is 0 or 2, respectively. Either way, we obtain a value of $1/2^m = \mu(X_{i_0\cdots i_{m-1}})$, thus proving invariance. □

References

[1] Michael F. Barnsley and John H. Elton. A new class of Markov processes for image encoding. *Advances in Applied Probability*, 20(1):14–32, 1988.

[2] Yuan Shih Chow and Henry Teicher. *Probability Theory. Independence, Interchangeability, Martingales.* Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1988.

[3] Andrea Crisanti, Angelo Vulpiani, and Giovanni Paladin. *Products of Random Matrices in Statistical Physics*, volume 104 of *Springer Series in Solid-State Sciences.* Springer-Verlag, Berlin, 1993.

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

[5] John Elton. An ergodic theorem for iterated maps. *Ergodic Theory and Dynamical Systems*, 7:481–488, 1987.

[6] Gary Froyland and Kazuyuki Aihara. Rigorous numerical estimation of Lyapunov exponents and invariant measures of iterated function systems and random matrix products. Submitted to *International Journal of Bifurcation and Chaos*.

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

[8] Kazushige Ishii. Localization of eigenstates and transport phenomena in the one-dimensional disordered system. *Supplement of the Progress of Theoretical Physics*, 53:77–138, 1973.

[9] J.E.Hutchinson. Fractals and self similarity. *Indiana University Mathematics Journal*, 30:713–747, 1981.

[10] Richard Kenyon and Yuval Peres. Intersecting random translates of invariant Cantor sets. *Inventiones mathematicae*, 104:601–629, 1991.

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

[12] Taijiro Ohno. Asymptotic behaviors of dynamical systems with random parameters. *Publications of the Research Institute for Mathematical Sciences, Kyoto University*, 19:83–98, 1983.

[13] Yuval Peres. Domains of analytic continuation for the top Lyapunov exponent. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, 28(1):131–148, 1992.

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

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