Pergamon

Nonlinear Analysis, Theory, Methods & Applications, Vol. 32, No. 7, pp. 831-860, 1998

Printed in Great Britain
0362-546X/98 $19.00+0.00

PII: S0362-546X(97)00527-0

APPROXIMATING PHYSICAL INVARIANT MEASURES OF MIXING DYNAMICAL SYSTEMS IN HIGHER DIMENSIONS

GARY FROYLAND†

Department of Mathematics, The University of Western Australia, Nedlands, WA 6907, Australia

(Received 6 September 1996; received in revised form 14 January 1997; received for publication 1997)

Key words and phrases: Invariant measure, mixing dynamical systems, Perron–Frobenius operator, Ulam approximation.

1. INTRODUCTION

In dynamical systems, one is often interested in the statistical properties of orbits of some dynamical system (M, T), where $\mathfrak{T}: M \circlearrowleft$, and typically $M \subset \mathbb{R}^d$ is some smooth manifold. For many systems it appears, at least in computer simulations, that there is a special distribution that nearly all orbits of T exhibit. In other words, if one plots the orbit $\{T^i x\}_{i=0}^N$ (N large), one obtains roughly the same picture for nearly all starting points $x \in M$ (excepting orbits beginning on a periodic point, for example). The invariant measures of T describe the asymptotic behavior of orbits of T for various starting points in the following sense. Suppose that $g: M \to \mathbb{R}$ is some test function that we wish to average along an orbit of T beginning at x. That is, we want to compute the time average

$$\mathfrak{T}_g(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i x). \tag{1}$$

For any ergodic T-invariant probability measure μ, the Birkhoff ergodic theorem says that the time average $\mathfrak{T}_g(x)$ of g along the orbit of x under T, is μ-almost everywhere equal to the space average $\mathfrak{S}_g$ of g, where

$$\mathfrak{S}_g = \int_M g(x)\, \mathrm{d}\mu(x). \tag{2}$$

Substituting $g = \chi_E$, where χ_E is the characteristic function of some measurable set $E \subset M$, we obtain

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \mathrm{card}\{i \in [0, n-1] : T^i x \in E\} &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_E(T^i x) \\ &= \int_M \chi_E(x)\, \mathrm{d}\mu(x) = \mu(E),\end{aligned} \tag{3}$$

for μ-almost all $x \in M$. Thus $\mu(E)$ tells us the fraction of time an orbit of x spends in $E \subset M$ for μ-almost all x. If it so happens that μ has very small support (for example μ has support on a periodic orbit), then the Birkhoff theorem does not give us any information about orbits not starting in this small set. Clearly the amount of information we get depends very much on our choice of μ. We would like to use the invariant measure which

† Email: gary@sat.t.u-tokyo.ac.jp; URL: http//www.sat.u-tokyo.ac.jp/~gary.

is exhibited in computer simulations by almost all "randomly" chosen starting points. For certain systems, including transitive expanding maps and many Anosov diffeomorphisms there is a unique ergodic invariant measure μ_{abs} that is equivalent to Lebesgue measure. Such measures guarantee that time averages $\mathfrak{T}_g(x)$ coincide with space averages $\mathcal{S}_g$ for *Lebesgue*-almost all starting points $x \in M$. Clearly this is a strong statement; the measure μ_{abs} may be considered to be the "physical" or "natural" measure of the dynamical system as it describes the distribution of orbits for Lebesgue-almost all starting points. It is the absolutely continuous invariant measure that we wish to approximate.

Let h be the density function for μ_{abs}, that is $\mu_{\mathrm{abs}}(E) = \int_E h\,dm$, where m is normalised Lebesgue measure on M. It is known [1] that h is a fixed point of the Perron–Frobenius operator of T, $\mathcal{P} : L^1(M, m) \circlearrowleft$ defined by

$$\mathcal{P}f(x) = \sum_{y \in T^{-1}x} \frac{f(y)}{|\det DT(y)|}, \tag{4}$$

or equivalently,

$$\int_A \mathcal{P}f\,dm = \int_{T^{-1}A} f\,dm \quad \text{for all Lebesgue measurable } A \subset M. \tag{5}$$

As $\mathcal{P}$ is an infinite-dimensional linear operator, the eigenvalue equation $h = \mathcal{P}h$ is not easily solvable analytically. A finite dimensional approximation to $\mathcal{P}$, originally put forward by Ulam [2], is constructed from a finite (measurable) partition of the state space. Let $\mathfrak{B} = \{A_1, \ldots, A_n\}$ be a finite partition of M with each A_i connected, and construct the $n \times n$ matrix $\tilde{P}_n$ as

$$\tilde{P}_{n,ij} = \frac{m(A_i \cap T^{-1}A_j)}{m(A_i)}. \tag{6}$$

$\tilde{P}_n$ is a stochastic matrix and the (assumed unique) left eigenvector $\tilde{p}_n$ of unit eigenvalue ($\tilde{p}_n\tilde{P}_n = \tilde{p}_n$) defines a measure on M by

$$\tilde{\mu}_n(E) := \sum_{i=1}^{n} \frac{m(E \cap A_i)}{m(A_i)} \cdot \tilde{p}_{n,i}. \tag{7}$$

Note that $\tilde{\mu}_n(A_i) = \tilde{p}_{n,i}$ for all $1 \le i \le n$, and that weight within each partition set A_i is distributed uniformly. One refines the partition sets so that the maximum diameter of the sets goes to zero and extracts a limiting measure $\tilde{\mu}_\infty$ from the sequence $\{\tilde{\mu}_n\}_{n=n_0}^{\infty}$, that is, $\tilde{\mu}_\infty = \lim_{n\to\infty} \tilde{\mu}_n$. (For ease of notation, the nth partition $\mathfrak{B}_n$ contains n sets. This need not be the case; but relabelling the index set for the partitions just complicates notation.) It is known [3, 4] that $\tilde{\mu}_\infty$ is T-invariant, but to identify whether or not $\tilde{\mu}_\infty$ is the "physical" invariant measure is much more difficult. The most complete result to date for piecewise C^2 expanding multidimensional mappings is that of Ding and Zhou [5, 6], who show that the sequence μ_n (defined by a slightly more complicated $\tilde{P}$) does indeed tend to the unique absolutely continuous invariant measure μ_{abs}. This paper seeks an extension of this result to the class of mixing *hyperbolic* mappings on M that possess a unique invariant measure equivalent to Lebesgue. For two-dimensional Anosov systems, this result has been shown to be true [7] using techniques such as symbolic dynamics and equilibrium states, provided that a special partition known as a Markov partition is used. This is restrictive for computer experiments, and here we look at using arbitrary (connected, measurable, and regular) partitions. Our method of attack is completely different from that of [7].

2. DYNAMICAL SYSTEMS CONSIDERED

In this paper we aim to generalise the main result of [7] by relaxing the requirement that the partitions $\mathfrak{B}_n$ be Markov partitions. As soon as we throw away the Markov property, we are no longer able to code our dynamical system as a subshift of finite type. In addition, the entries of the matrix P no longer have the lovely interpretation as an approximation to the rate of stretching in unstable directions. We are back to square one, and must devise a completely new method of proof that does not rely on coding at all.

We will be mainly concerned with $C^{1+\mathrm{Lip}}$ Anosov systems, however, the multidimensional expanding case is also covered. The properties of T that we are after are:

(i) T possesses a unique absolutely continuous invariant measure μ with density h.

(ii) T is mixing with respect to μ.

(iii) h is strictly positive

(iv) h is Lipschitz.

Theorem III.1.1 of Mañé [8] guarantees that these four properties are satisfied by our class of expanding maps. We work with the class of transitive $C^{1+\mathrm{Lip}}$ Anosov maps that possess an absolutely continuous invariant measure μ; Bowen [9] Theorem 4.14 and Corollary 4.13 then guarantee that properties (ii)–(iv) are satisfied. The main result of this paper also holds for $C^{1+\gamma}$ maps, $0 < \gamma < 1$ (with corrections to the rate of convergence), however for clarity of presentation we restirct ourselves to the Lipschitz derivative case.

3. OUTLINE OF METHOD

We first list our assumptions for the class of partitions that are considered.

Assumptions 3.1. Let $\{\mathfrak{B}_n\} = \{\{A_1^n, \ldots, A_n^n\}\}_{n=1}^{\infty}$ be a sequence of partitions with the maximum diameter of the partition sets going to zero as $n \to \infty$. Standing hypotheses are that

(i) the elements of $\mathfrak{B}_n$ are connected and Lebesgue measurable with positive measure for every $n = 1, 2, \ldots$.

(ii) there is a constant $\alpha < \infty$ such that a d-dimensional hypercube with edge length $D_n = \max_{1 \le i \le n} A_i^n$ may cover at most α of the sets $\{A_1^n, \ldots, A_n^n\}$ for each $n = 1, 2, \ldots$.

(iii) there is a constant $\beta < \infty$ such that $\max_{1 \le i,j \le n} m(A_i^n)/m(A_j^n) \le \beta$ for $n = 1, 2, \ldots$.

Assumption 3.1(ii) is a geometric condition that ensures that our partition sets are roughly the same shape and not too thin or spiky. It also guarantees that the partition sets are spread out though m in a roughly uniform manner, without large numbers of sets occupying small regions of M. Assumption 3.1(iii) just says that the partition sets are all roughly the same size. These hypotheses are very mild and are there to rule out any pathological cases. Almost any sequence of partitions that one is likely to use for computational purposes with satisfy these conditions.

Example 3.2. Suppose we have a simple partition of M into n cubes in a standard grid-like formation. Clearly Assumption 3.1(i) is satisfied; we show (ii). Each partition set has the same diameter, say D_n, and a hypercube of diameter D_n may not cover more than 2^d cubes, where d is the dimension of M. Thus $\alpha = 2^d$ in this case. Finally (iii) holds with $\beta = 1$.

We now state our main result.

THEOREM 3.3. Let $\{\tilde{P}_n\}_{n=n_0}^{\infty}$ be a sequence of $(n \times n)$ transition matrices generated by (6) from a sequence of partitions $\{\mathfrak{B}_n\}_{n=n_0}^{\infty}$ satisfying Assumption 3.1 whose maximal element diameter goes to zero as $n \to \infty$. Define $\tilde{R}_n$ and $\tilde{r}_n$ to be the least values satisfying

$$\max_{1 \le i \le n} \sum_{j=1}^{n} |\tilde{P}_{ij}^k - \tilde{p}_j| \le \tilde{R}_n \tilde{r}_n^k \qquad \text{for all } n, k \ge 0.$$

If

(i) $\tilde{r}_n \le \tilde{r}'$, for all $n \ge n_0$ and some $\tilde{r}' < 1$, and
(ii) $\tilde{R}_n = O(n^p)$, $p > 0$

then the sequence of invariant measures $\{\tilde{\mu}_n\}_{n=n_0}^{\infty}$ defined by (7) converges strongly to the unique absolutely continuous measure μ of T. The rate of convergence is $O(\log n/n^{1/d})$.

In later sections, we will study two classes of maps for which we are able to verify the conditions (i) and (ii). Numerical results are also detailed for a common multidimensional map, providing evidence that one can expect conditions (i) and (ii) to hold for mixing higher dimensional systems. The remainder of this section sets up our plan of attack for the proof of Theorem 3.3.

Observation 3.4. If we were to define a new matrix P by

$$P_{ij} = \frac{\mu(A_i \cap T^{-1}A_j)}{\mu(A_i)} \tag{8}$$

then the vector $p = [\mu(A_1), \mu(A_2), \ldots, \mu(A_n)]$ would be the left eigenvector of P of unit eigenvalue due to the invariance of μ.

A simple calculation shows that $p_j = \sum_{i=1}^{n} p_i P_{ij}$.

$$\begin{aligned}\sum_{i=1}^{n} p_i P_{ij} &= \sum_{i=1}^{n} \mu(A_i) \cdot \frac{\mu(A_i \cap T^{-1}A_j)}{\mu(A_i)} \\ &= \sum_{i=1}^{n} \mu(A_i \cap T^{-1}A_j) \\ &= \mu(T^{-1}A_j) \\ &= \mu(A_j) \quad \text{by } T\text{-invariance of } \mu \\ &= p_j.\end{aligned}$$

So if we knew beforehand what μ was, we could use (8) to calculate the *exact* value of $\mu(A_i)$ for each $i = 1, \ldots, n$. This observation alone seems of little value as our problem is to estimate μ.

Observation 3.5. The elements of $\tilde{P}$ are not very different from P.

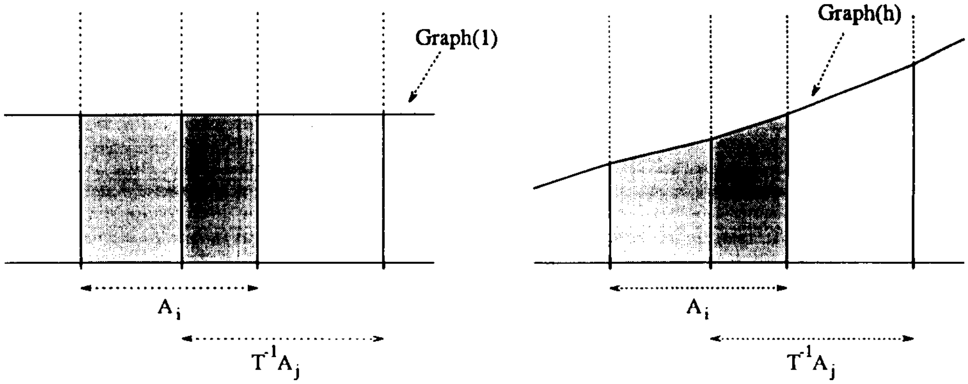

Fig. 1. The plot on the left is the graph of the density function for Lebesgue measure and the one on the right is the graph of the density function of the T-invariant measure μ.

This requires some thought, but is due to the fact that μ and m are *equivalent* measures. It is easy to see in one dimension. In Fig. 1, the density on left is identically equal to unity (Lebesgue measure) and the density on the right is the density for our absolutely continuous invariant measure μ. Because the density h is Lipschitz and bounded away from zero, the fraction of the heavily shaded region contained in the shaded region is roughly the same in both pictures. As the length of the subintervals A_i and $T^{-1}A_j$ decreases, these ratios converge to the same value. More precisely, we have the following lemma.

LEMMA 3.6. Let $\{\mathfrak{B}_n\}_{n=n_0}^{\infty} = \{\{A_1^n, A_2^n, \ldots, A_n^n\}\}_{n=n_0}^{\infty}$ be a sequence of partitions of a compact set $M \subset \mathbb{R}^d$ satisfying Assumptions 3.1. Let P_n and $\tilde{P}_n$ be the matrices obtained from these partitions as above. Then $|\tilde{P}_{n,ij} - P_{n,ij}| = O(n^{-1/d})$.

Proof. Suppose first that $m(A_i^n \cap T^{-1}A_j^n) = 0$. Then one has $\mu(A_i^n \cap T^{-1}A_j^n) = 0$ also as $\mu \ll m$, and the difference $\tilde{P}_{n,ij} - P_{n,ij} = 0$. In the situation where $m(A_i^n \cap T^{-1}A_j^n) \neq 0$, we proceed as follows.

$$
\begin{aligned}
|\tilde{P}_{n,ij} - P_{n,ij}| &= \left| \frac{m(A_i^n \cap T^{-1}A_j^n)}{m(A_i^n)} - \frac{\mu(A_i^n \cap T^{-1}A_j^n)}{\mu(A_i^n)} \right| \\
&= \frac{m(A_i^n \cap T^{-1}A_j^n)}{m(A_i^n)} \left| 1 - \frac{\mu(A_i^n \cap T^{-1}A_j^n)}{m(A_i^n \cap T^{-1}A_j^n)} \cdot \frac{m(A_i^n)}{\mu(A_i^n)} \right| \\
&= \frac{m(A_i^n \cap T^{-1}A_j^n)}{m(A_i^n)} \left| 1 - \frac{1/m(A_i^n \cap T^{-1}A_j^n) \int_{A_i^n \cap T^{-1}A_j^n} h \, \mathrm{d}m}{1/m(A_i^n) \int_{A_i^n} h \, \mathrm{d}m} \right| \\
&\leq \frac{m(A_i^n \cap T^{-1}A_j^n)}{m(A_i^n} \left| 1 - \left(\sup_{x \in A_i^n \cap T^{-1}A_j^n} h(x) \right) \cdot \left(\inf_{x \in A_i^n} h(x) \right)^{-1} \right| \\
&\leq 1 \cdot \left(\inf_{x \in A_i^n} h(x) \right)^{-1} \left| \sup_{x \in A_i^n \cap T^{-1}A_j^n} h(x) - \inf_{x \in A_i^n} h(x) \right| \\
&\leq \left(\inf_{x \in A_i^n} h(x) \right)^{-1} \left| \sup_{x \in A_i^n} h(x) - \inf_{x \in A_i^n} h(x) \right| \\
&\leq \left(\inf_{x \in M} h(x) \right)^{-1} \mathrm{Lip}(h) \max_{1 \leq i \leq n} \mathrm{diam}(A_i^n),
\end{aligned}
$$

where Lip(h) is the Lipschitz constant of h. We now wish to bound

$$D_n = \max_{1 \le i \le n} \operatorname{diam}(A_i^n)$$

in terms of the number of partition sets n. Put $q_n = [\operatorname{diam}(M)/D_n] + 1$, where $[\cdot]$ denotes integer part. We may cover M with q_n^d cubes of side length D_n.† Thus the number of partition sets n is bounded by $q_n^d \alpha$. So

$$n \le q_n^d \alpha \Leftrightarrow \left(\frac{n}{\alpha}\right)^{1/d} \le q_n \le \frac{\operatorname{diam}(M)}{D_n} + 1$$

$$\Rightarrow \frac{\operatorname{diam}(M)}{(n/\alpha)^{1/d} - 1} \ge D_n,$$

and we have

$$|\tilde{P}_{n,ij} - P_{nij}| \le \frac{\operatorname{Lip}(h) \operatorname{diam}(M)}{((n/\alpha)^{1/d} - 1) \inf h} \tag{9}$$

$$\le \frac{C \operatorname{Lip}(h) \operatorname{diam}(M) \alpha^{1/d}}{\inf h} \tag{10}$$

for some constant $C > 0$. ■

We are aiming for a bound on the matrix 1-norm of the error matrix $E_n := \tilde{P}_n - P_n$. This norm is induced by the standard L^1 vector norm under left multiplication and $\|E_n\|_1$ is equal to the maximal row sum of $|\tilde{P}_n - P_n|_{ij}$. The following two observations help us in this direction.

Observation 3.7. Since μ and m are equivalent measures, the matrix $E_n = \tilde{P}_n - P_n$ will only have nonzero entries where P_n (and $\tilde{P}_n$) are nonzero.

Observation 3.8. There is a universal bound for the number of nonzero entries in each row of P_n, $n = 1, 2, \ldots$, depending only on α and the Lipschitz constant λ of T.

To see this, fix i and count the number of nonzero entries in the ith row of P. If $m(A_i \cap T^{-1}A_j) > 0$, then $T(A_i) \cap A_j \ne \emptyset$. So for a fixed i, how many sets A_j can $T(A_i)$ intersect? The diameter of $T(A_i)$ is bounded by $\lambda \operatorname{diam}(A_i)$, and so we may bound the number of A_j's that can be intersected by $T(A_i)$ using α; this will be shown in detail in the following proof. We now put Observations 3.7 and 3.8 and Lemma 3.6 together to bound $\|E_n\|_1$.

LEMMA 3.9. *Under the hypotheses of Lemma 3.6,* $\|E_n\|_1 = O(n^{-1/d})$.

---

† The fact that a set of diameter one may be covered by a hypercube of edge length one is a result of a Helly-type theorem of Klee (Theorem 6.5 in Lay [10] for example). This result reduces the proof to showing that a d-dimensional simplex of diameter one may be covered by a d-dimensional hypercube of edge length one.

Proof. All we need to show is that the number of nonzero entries in any row of E_n is bounded by some constant value for all $n \geq 0$. The result then follows from Lemma 3.6. Note that $T(A_i^n \cap T^{-1}A_j^n) = T(A_i^n) \cap A_j^n$, so that if $\mu(A_i^n \cap T^{-1}A_j^n) > 0$, then $A_i^n \cap T^{-1}A_j^n \neq \varnothing$ and so too must $T(A_i^n) \cap A_j^n \neq \varnothing$. So if we count up how many times $TA_i^n \cap A_j^n \neq \varnothing$, for a fixed i, this will be an upper bound for the number of nonzero entries in row i of E_n.

The maximum number of sets that the image of any single set can touch is no greater than

$$([\|DT\|] + 1)^d \cdot \alpha \tag{11}$$

where $\|DT\| := \sup_{x \in M}\|D_x T\|$. This is because the diameter of any set of the form $T(A_i^n)$ is bounded above by $\|DT\| \cdot \max_{1 \leq i \leq n}\operatorname{diam}(A_i^n)$. We may cover a set with diameter $\|DT\| \cdot D_n$ by $([\|DT\|] + 1)^d$ cubes of edge length D_n, and the bound (11) follows. So using Lemma 3.6 we finally have

$$\begin{aligned}\|E_n\|_1 &\leq ((\|DT\| + 1)^d\alpha)\left(\frac{C\operatorname{Lip}(h)\operatorname{diam}(M)\alpha^{1/d}}{\inf h}n^{-1/d}\right)\\ &= \frac{C(\|DT\| + 1)^d\alpha^{1+1/d}\operatorname{diam}(M)\operatorname{Lip}(h)}{\inf h}\cdot\frac{1}{n^{1/d}}. \qquad \blacksquare\end{aligned} \tag{12}$$

At this stage, we know that $\tilde{P}_n$ and P_n are close in the matrix 1-norm sense, and have an estimate of their rate of convergence. Since $\tilde{P}_n$ and P_n are close, so too are $\tilde{p}_n$ and p_n. In the next section we see that if $\|\tilde{p}_n - p_n\|_1 \to 0$ then $\tilde{\mu}_n \to \mu$ strongly, where $\tilde{\mu}_n$ is our approximate invariant measure given in (7).

4. L^1 AND STRONG CONVERGENCE

Lasota and Mackey [1, p. 402] show that strong convergence† of a sequence of absolutely continuous probability measures μ_n to an absolutely continuous probability measure μ is equivalent to strong convergence (L^1-norm convergence) of the density function ϕ_n of μ_n to the density function h of μ.

Our plan of action is as follows. To show that $\|\tilde{\mu}_n - \mu\|$ goes to zero as $n \to \infty$, ($\|\cdot\|$ is the norm for the strong operator topology), we show separately that $\|\tilde{\mu}_n - \mu_n\|$ and $\|\mu_n - \mu\|$ go to zero.

LEMMA 4.1. The following are equivalent.

(i) $\|p_n - \tilde{p}_n\|_1 \to 0$ as $n \to \infty$, where p_n and $\tilde{p}_n$ are vectors of length n.

(ii) $\tilde{\mu}_n \to \mu_n$ strongly where μ_n is the measure defined by

$$\mu_n(E) = \sum_{i=1}^n \frac{m(E \cap A_i^n)}{m(A_i^n)} \cdot p_{n,i} \qquad \text{and} \qquad \tilde{\mu}_n(E) = \sum_{i=1}^n \frac{m(E \cap A_i^n)}{m(A_i^n)} \cdot \tilde{p}_{n,i}.$$

†The definition of strong convergence of measures in [1] is equivalent to norm convergence in $(C(M, \mathbb{R}))^*$, the space of all continuous linear functionals on $C(M, \mathbb{R})$; see, for example, Conway [11, Appendix C]. Since $(C(M, \mathbb{R}))^*$ may be identified with the space of all regular Borel measures on M, strong convergence as defined in [1] is strong convergence in the usual sense of norm convergence on $(C(M, \mathbb{R}))^*$.

Proof. We have the associated densities for μ_n and $\tilde{\mu}_n$.

$$\phi_n(x) = \sum_{i=1}^{n} \frac{p_{n,i}}{m(A_i^n)} \chi_{A_i^n}(x) \quad \text{and} \quad \tilde{\phi}_n(x) = \sum_{i=1}^{n} \frac{\tilde{p}_{n,i}}{m(A_i^n)} \chi_{A_i^n}(x).$$

$$\begin{aligned}
\|\phi_n - \tilde{\phi}_n\|_1 &= \int_M |\phi_n - \tilde{\phi}_n| \, \mathrm{d}m \\
&= \int_M \left| \sum_{i=1}^{n} \frac{p_{n,i} - \tilde{p}_{n,i}}{m(A_i^n)} \chi_{A_i^n} \right| \mathrm{d}m \\
&= \int_M \sum_{i=1}^{n} \frac{|p_{n,i} - \tilde{p}_{n,i}|}{m(A_i^n)} \chi_{A_i^n} \, \mathrm{d}m \quad \text{as we have a sum over disjoint sets,} \\
&= \sum_{i=1}^{n} |p_{n,i} - \tilde{p}_{n,i}| \\
&= \|p_n - \tilde{p}_n\|_1.
\end{aligned}$$

Using the result in [1], we are done. ∎

LEMMA 4.2. Define $\mu_n : \mathfrak{B}(M) \to \mathbb{R}^+$ by $\mu_n(E) = \sum_{i=1}^{n} m(E \cap A_i^n)/m(A_i^n) \cdot \mu(A_i^n)$. If μ is absolutely continuous, then μ_n converges strongly to μ as $n \to \infty$.

Proof. The density of μ_n is $\phi_n(x) = \sum_{i=1}^{n} (\int_{A_i^n} h \, \mathrm{d}m)/m(A_i^n) \cdot \chi_{A_i^n}(x)$, where h is the density of μ. Note that ϕ_n is really just a discretised version of h, taking on the average values of h over each partition set. The proof that $\phi_n \to h$ strongly goes through as in [12, Lemma 2.2] and the result follows. ∎

Finally, we have the following.

PROPOSITION 4.3. Let p_n be the unique positive left eigenvector of P_n, the $n \times n$ transition matrix generated using μ, and $\tilde{p}_n$ be the corresponding eigenvector of $\tilde{P}_n$, the transition matrix generated using Lebesgue measure. Let $\tilde{\mu}_n : \mathfrak{B}(M) \to \mathbb{R}^+$ be defined by

$$\tilde{\mu}_n(E) = \sum_{i=1}^{n} \frac{m(E \cap A_i^n)}{m(A_i^n)} \cdot \tilde{p}_{n,i}.$$

If $\|p_n - \tilde{p}_n\|_1 \to 0$ as $n \to \infty$, then $\tilde{\mu}_n$ converges strongly to μ.

Proof. Immediate from Lemmas 4.1 and 4.2. ∎

We now have something to aim for, namely to show that $\|\tilde{p}_n - p_n\|_1 \to 0$ as $n \to \infty$. But does norm convergence of $\tilde{P}_n$ to P_n tell us that $\|\tilde{p}_n - p_n\|_1 \to 0$? It is not obvious that this is so, as $\tilde{P}_n$ and P_n are $n \times n$ matrices and grow in size as n increases. To try to answer this, we treat the stochastic matrices $\tilde{P}_n$ as transition matrices of finite state Markov chains, and consider the *sensitivity* of these chains to perturbations of their transition probabilities.

5. SENSITIVITY OF FINITE MARKOV CHAINS

In this section, we drop the n dependence on the matrices P_n, and just consider a fixed transition matrix. The sensitivity of a finite Markov chain is a measure of how much the invariant density changes in response to a perturbation in the elements of the transition matrix. An important equality concerning this sensitivity due to Schweitzer [13] (see also Haviv and Van der Heyden [14] and Seneta [15]) is,

$$\tilde{p} - p = \tilde{p}(\tilde{P} - P)(I - P + P^{\infty})^{-1} \tag{13}$$

where $\tilde{p}$ and p are the normalized left eigenvectors associated with the eigenvalue one for $\tilde{P}$ and P, respectively. P^{∞} denotes the transition matrix that has p as its rows. Equation (13) may be checked using simple matrix arithmetic. The matrix $Z := (I - P + P^{\infty})^{-1}$ is the *fundamental matrix* of the Markov chain governed by P; see Kemeny and Snell [16]. This matrix exists as $P - P^{\infty}$ has spectral radius strictly less than unity, and so Z may be represented as

$$Z = (I - (P - P^{\infty}))^{-1} = \sum_{k=0}^{\infty} (P - P^{\infty})^k. \tag{14}$$

From (13) we immediately get the inequality,

$$\|\tilde{p} - p\|_1 \leq \|\tilde{P} - P\|_1 \cdot \|Z\|_1. \tag{15}$$

Recall that the matrix 1-norm is the maximum row sum of the absolute values of the entries as we are doing left multiplication (normally it is the maximum column sum). We have seen that 1-norm convergence of $\tilde{p}_n$ to p_n is equivalent to strong convergence of $\tilde{\mu}_n$ to μ, where $\tilde{\mu}_n$ is the measure defined by (7), and μ is the absolutely continuous measure. Since we know the rate at which the first term of the right-hand side of (15) goes to zero, our remaining task is to prove that the second term grows strictly slower than $O(n^{1/d})$. Our aim is to show that $\|Z_n\| = O(\log n)$, where Z_n is the fundamental matrix for the Markov chain with transition matrix P_n, as this will guarantee convergence of our approximate measure $\tilde{\mu}_n$ to the absolutely continuous measure μ in any dimension d. In fact the remainder of this article is devoted entirely to this difficult objective. A proof of such a result for Anosov maps is still lacking. In later sections, we discuss classes of maps for which we are able to obtain rigorous convergence results. For more general maps we provide heuristic arguments and cite numerical results to support our conjecture of logarithmic growth of Z_n.

6. FACTORS INFLUENCING $\|Z_n\|_1$

Roughly speaking, the norm of Z_n is a measure of how rapidly initial distributions approach equilibrium.

Definition 6.1. A Markov chain with transition matrix P is *uniformly ergodic* if

$$\max_{1 \leq i \leq n} \|P^k(i, \cdot) - p(\cdot)\|_1 \to 0, \qquad \text{as } k \to \infty. \tag{16}$$

Remark 6.2. We will interchangeably use the notation P_{ij} and $P(i, j)$ to represent the i, jth element of the matrix P.

As $\mu \equiv m$, it is easy to show that

(i) (T, μ) ergodic $\Rightarrow$ P irreducible,

(ii) (T, μ) mixing $\Rightarrow$ P irreducible and aperiodic.

Irreducibility and aperiodicity make P ergodic, and the notion of uniform ergodicity and ergodicity coincide for finite state Markov chains. It is known that finite state ergodic Markov chains enjoy the property of exponential mixing (see [17] for example).

THEOREM 6.3. If a Markov chain with transition matrix P is uniformly ergodic, then there exists constants $R < \infty$, $0 < r < 1$ such that

$$\max_{1 \le i \le n} \|P^k(i, \cdot) - p(\cdot)\|_1 \le Rr^k. \tag{17}$$

The least such constant r is referred to as the *rate of mixing* of the Markov chain, and is equal to the second largest (in absolute value) eigenvalue of P. To see that the norm of $Z = Z(P)$ is related to the rate of mixing of P, we note the following result, whose proof may be found in Kemeny and Snell [16].

THEOREM 6.4. For any ergodic Markov chain with transition matrix P,

$$Z = 1 + \sum_{k=1}^{\infty} (P^k - P^\infty). \tag{18}$$

By using the bound

$$\|Z\|_1 \le 1 + \sum_{k=1}^{\infty} \|P^k - P^\infty\|_1 \le 1 + \sum_{k=1}^{\infty} Rr^k \le 1 + \frac{Rr}{1-r}, \tag{19}$$

we see how the rate of mixing r may influence the norm of Z. Recall that we are interested in the growth of the norm of $Z_n := Z(P_n)$ with n. For each n, Theorem 6.3 guarantees the existence of constants R_n and r_n such that (17) is satisfied with these constants, upon insertion of P_n and p_n. We continue under the assumption that r_n may be universally bounded away from unity, that is, there exists $r' < 1$ such that $r_n \le r'$ for all $n \ge n_0$. This seems plausible as the dynamical systems we are dealing with are exponentially mixing with respect to the "physical" measure. This matter will be discussed later. For the moment, we consider the growth of R_n.

We require the following lemma, which is a special case of Theorem 16.2.4 [17].

LEMMA 6.5 (Meyn and Tweedie [17]). Let P be an ergodic (irreducible, aperiodic) $n \times n$ transition matrix and p its unique invariant density. Set m to be the least nonnegative integer such that

$$(P^m)_{ij} \ge \frac{p_j}{2} \qquad \text{for all } 1 \le i, j \le n. \tag{20}$$

Then

$$\max_{1 \le i \le n} \|P^k(i, \cdot) - p(\cdot)\|_1 \le \left(\frac{1}{2}\right)^{k/m} \qquad \text{for all } k = 0, 1, \ldots. \tag{21}$$

This lemma gives a bound for the asymptotic rate of convergence of the Markov chain governed by P to equilibrium. The only information that is used is the time (m steps) that it takes an initial mass concentrated in a single state i to be spread to within "half of equilibrium" at j. It is important to note that the estimate here does not involve a constant as in (17). The hypothesis of the lemma may seem a little unusual, and to remedy this, we have the following alternative, with the hypothesis stated in terms of the matrix 1-norm.

LEMMA 6.6. For any uniformly ergodic Markov chain with transition matrix P and invariant density p, if

$$\max_{1 \le i \le n} \|P^m(i, \cdot) - p(\cdot)\|_1 \le \frac{\min_{1 \le i \le n} p_i}{2} \qquad \text{for some } m \ge 0, \tag{22}$$

then

$$\max_{1 \le i \le n} \|P^k(i, \cdot) - p(\cdot)\|_1 \le \left(\frac{1}{2}\right)^{k/m}, \qquad \text{for all } k \ge 0.$$

Proof. This is clear because

$$|P_{ij}^m - p_j| \le \max_{1 \le i \le n} \|P^m(i, \cdot) - p(\cdot)\|_1 \le \frac{\min_{1 \le i \le n} p_i}{2} \le \frac{p_j}{2},$$

and $|P_{ij}^m - p_j| \le p_j/2$ implies $P_{ij}^m \ge p_j/2$. The result now follows from Lemma 6.5. ∎

At this stage, we note that if $\|P^k - P^\infty\|_1 \le (\frac{1}{2})^{k/m}$, then

$$\|Z\|_1 \le \frac{1}{1 - (\frac{1}{2})^{1/m}}, \tag{23}$$

using the first inequality of (19). We need the following easily proven facts.

LEMMA 6.7.

(i)

$$\left|\frac{1}{1 - a^{1/bx}} - \left(\frac{bx}{\log(1/a)} + \frac{1}{2}\right)\right| \to 0 \qquad \text{as } x \to \infty,\ 0 < a < 1,\ b \ne 0 \tag{24}$$

(ii)

$$\frac{1}{1 - a^{1/bx}} \le \frac{bx}{\log(1/a)} + 1 \qquad x \ge 0,\ 0 < a < 1,\ b \ne 0. \tag{25}$$

Let m_n be the minimal integer m appearing in either Lemma 6.5 or 6.6 corresponding to the transition matrix P_n. We now have from (23) that

$$\|Z_n\|_1 \le \frac{1}{1 - (\frac{1}{2})^{1/m_n}} \le \frac{m_n}{\log 2} + 1,$$

for all $n \ge 0$ and so

$$\|Z_n\|_1 = O(m_n). \tag{26}$$

Thus we have reduced the problem to showing that the integer m_n in either Lemma 6.5 or 6.6 grows logarithmically. As this number seems difficult to get one's hands on, we have the following, perhaps more direct characterisation, of convergence of the invariant densities (which is a modified restatement of Theorem 3.3).

PROPOSITION 6.8. Let $\{\tilde{P}_n\}_{n=n_0}^{\infty}$ be a sequence of $(n \times n)$ transition matrices generated by (6) from a sequence of partitions $\{\mathfrak{B}_n\}_{n=n_0}^{\infty}$ satisfying Assumption 3.1 whose maximal element diameter goes to zero as $n \to \infty$. If the constants R_n, r_n corresponding to P_n (as in Theorem 6.3) satisfy

(i) $r_n \le r'$, for all $n \ge n_0$ and some $r' < 1$, and

(ii) $R_n = O(n^p)$, $p > 0$

then the sequence of invariant measures $\{\tilde{\mu}_n\}_{n=n_0}^{\infty}$ defined by (7) converges strongly to the unique absolutely continuous measure μ of T. The rate of convergence is $O(\log n/n^{1/d})$.

Remark 6.9. Note that the constants R_n and r_n we are looking at belong to the matrix P_n and not $\tilde{P}_n$ as in Theorem 3.3. We choose to work with P_n for our proofs as it is the "special" finite approximation of the Perron-Frobenius operator whose left eigenvector p_n gives the *exact* weights for the partition sets A_i^n. We believe that these constants will be better behaved for P_n than for the rather arbitrary finite dimensional approximation $\tilde{P}_n$ using Lebesgue measure. Of course, if one is pursuing a numerical estimation of the behavior of these constants, then $\tilde{R}_n$ and $\tilde{r}_n$ are much more easily evaluated. The symmetry in equation (13) means that the above result holds for either pair of constants.

Proof of Proposition. We show that $\|Z_n\|_1 = O(\log n)$. This, combined with Lemma 3.9 and equation (15) will prove that $\|p_n - \tilde{p}_n\|_1 \to 0$. We then refer to Proposition 4.3 to see that this implies $\tilde{\mu}_n \to \mu$ strongly.

Fix n. We want to find the minimal m_n such that

$$R_n r_n^{m_n} \le \frac{\min_{1\le i\le n} p_{n,i}}{2}, \tag{27}$$

as this will allow us to apply Lemma 6.6 using m_n.

SUBLEMMA 6.10.

$$p_{n,i} = \mu(A_i^n) \ge \frac{\inf h}{\beta n} \qquad \text{for all } n \ge n_0 \text{ and } 1 \le i \le n.$$

Proof.

$$\begin{aligned}\min_{1\le i\le n} \mu(A_i^n) &\ge (\inf h) \min_{1\le i\le n} m(A_i^n)\\ &\ge (\inf h)\frac{\max_{1\le i\le n} m(A_i^n)}{\beta} \qquad \text{by Assumption 3.1(iii).}\\ &\ge (\inf h)\frac{m(M)}{\beta n} = \frac{\inf h}{\beta}\cdot\frac{1}{n}. \quad \blacksquare\end{aligned}$$

Let us find the minimal m_n satisfying

$$R_n r_n^{m_n} \le \frac{\inf h}{2\beta n}, \tag{28}$$

as such an m_n will also satisfy (27). Taking logs of both sides of (28), we obtain

$$\log R_n + m_n \log r_n \le \log\inf h - \log 2 - \log\beta - \log n,$$

and so

$$m_n \geq \frac{(\log R_n + \log n) + (\log 2 + \log \beta - \log \inf h)}{\log 1/r_n}.$$

We may therefore set

$$m_n = \left[\frac{(\log R_n + \log n) + (\log 2 + \log \beta - \log \inf h)}{\log 1/r'}\right] + 1,$$

and (27) will be satisfied. Here $[\cdot]$ denotes integer part, and note also, the replacement of r_n with r'. Thus $R_n = O(n^p)$, for fixed $p > 0$ will give us that $m_n = O(\log n)$ and by (26) we have that $\|Z_n\|_1 = O(\log n)$. By Lemma 3.9, we have $\|\tilde{p}_n - p_n\|_1 = O(\log n/n^{1/d})$.

It remains to show that $\|\mu - \tilde{\mu}_n\| = O(\log n/n^{1/d})$. Denote by ϕ_n, ϕ_n, the densities of $\mu_n, \tilde{\mu}_n$, respectively.

$$\begin{aligned}
\|\mu - \tilde{\mu}_n\| &= \|h - \tilde{\phi}_n\|_1 \qquad \text{see [1, p. 402]}\\
&\leq \|h - \phi_n\|_1 + \|\phi_n - \tilde{\phi}_n\|_1\\
&= \|h - \phi_n\|_1 + \|p_n - \tilde{p}_n\|_1 \qquad \text{as in the proof of Lemma 4.1}\\
&\leq \operatorname{Lip}(h) \cdot \left(\max_{1 \leq i \leq n} \operatorname{diam}(A_i^n)\right) + O(\log n/n^{1/d})\\
&\leq \operatorname{Lip}(h) \cdot \frac{C\alpha^{1/d} \operatorname{diam}(m)}{n^{1/d}} + O(\log n/n^{1/d}) \qquad \text{as in the proof of Lemma 3.6}\\
&= O(\log n/n^{1/d}). \quad \blacksquare
\end{aligned} \tag{29}$$

Remarks 6.11.

(i) Notice that if we were to blindly use the bound given by the right-hand side of (19), we would obtain the far worse estimate that $\|Z_n\|_1 = O(R_n)$, rather than $\|Z_n\|_1 = O(\log R_n)$. The bound of $\|Z_n\|_1 = O(R_n)$ would not give us convergence of $\tilde{p}_n$ to p_n, as we expect R_n to grow polynomially with n.

(ii) Bounds for $\|Z\|_1$ and/or $\max_{i,j}|z_{ij}|$ are considered in Meyer [18], and Seneta [15, 19]. The bound in [18] is far too conservative for our requirements; it shows that P having large diagonal elements and many eigenvalues close to one contribute to Z having large norm. To apply the results of [15], the matrix must be *scrambling* (have at least one positive column), so that the *ergodicity coefficient* is strictly less than unity; this is of no use to us as our matrices are far from scrambling. Finally, [19] bounds the norm of Z_n by its trace, but computer experiments show that the trace grows linearly with n; too fast for our purposes.

7. THE BEHAVIOUR OF R_n, r_n, AND Z_n FOR TWO CLASSES OF MAPS, AND NUMERICAL RESULTS

We have not yet addressed whether Assumptions (i) and (ii) in the preceding proposition are likely to be true. In the final three subsections, we consider two situations where we have good control over the behaviour of R_n and r_n and finish with some numerical results for more general maps.

7.1. Piecewise linear expanding Markov maps

In this section, we consider a particularly simple class of maps that allow us to have good control on the growth rate of the constants R_n and r_n. Our map $T\colon M\circlearrowleft$ is a piecewise linear expanding Markov map on some compact subset M of $\mathbb{R}^d$. More precisely, our map belongs to a class $\mathcal{K}$ described below.

Definition 7.1. A map $T\colon M\circlearrowleft$ belongs to *class* $\mathcal{K}$ if there exists a family of (Lebesgue) measurable, connected sets $\{A_1, \dots, A_n\}$ such that

(i) $\operatorname{Int} A_i \cap \operatorname{Int} A_j = \varnothing$ for $i \neq j$.

(ii) $\bigcup_i A_i = M$.

(iii) For each $1 \leq i \leq n$, there is a collection of indices $J_i = \{j_{i_1}, \dots, j_{i_r}\}$ such that $T(A_i) = \bigcup_{j\in J_i} A_j (\operatorname{mod} 0)$.

(iv) For every $x \in \bigcup_i \operatorname{Int} A_i$, the derivative of T exists and satisfies $\|D_x T\| > 1$. On each A_i, $D_x T$ is constant.

(v) T possesses a unique invariant measure equivalent to Lebesgue.

The family of measurable sets $\{A_1, \dots, A_n\}$ is referred to as a *Markov partition*. When $d = 1$, the partition sets are intervals, and when $d = 2$, we imagine the boundaries of these sets being formed by piecewise smooth lines running across M. We define a stochastic process $\{x_i : i = 0, 1, 2, \dots\}$ on the discrete space $S = \{1, \dots, n\}$, by

$$\operatorname{Prob}\{x_0 = i_0, x_1 = i_1, \dots, x_r = i_r\} = \mu(A_{i_0} \cap T^{-1}A_{i_1} \cdots \cap T^{-r}A_{i_r}),\quad i_k \in S,\ k = 0, 1, \dots, r.$$

This probability represents the measure of the set of points that lie in A_{i_1}, have their first iterate in A_{i_2}, their second iterate in A_{i_3}, and so on. Since μ is T-invariant, this stochastic process is stationary.

LEMMA 7.2. $\{x_i\}$, $i = 0, 1, 2, \dots$ is a Markov process.

Proof. For this to be true, the Markov property must hold; that is, we must have

$$\operatorname{Prob}\{x_2 = k \mid x_1 = j, x_0 = i\} = \operatorname{Prob}\{x_2 = k \mid x_1 = j\}, \quad \text{for all } 1 \leq i,j,k \leq n,$$

or in other words,

$$\operatorname{Prob}\{T^2x \in A_k \mid Tx \in A_j,\ x \in A_i\} = \operatorname{Prob}\{T^2x \in A_k \mid Tx \in A_j\}, \tag{30}$$

for all $x \in M$, and $1 \leq i,j,k \leq n$.

In terms of measures (30) reads

$$\frac{\mu(T^{-2}A_k \cap T^{-1}A_j \cap A_i)}{\mu(T^{-1}A_j \cap A_i)} = \frac{\mu(T^{-2}A_k \cap T^{-1}A_j)}{\mu(T^{-1}A_j)}. \tag{31}$$

Now

$$\frac{\mu(T^{-2}A_k \cap T^{-1}A_j)}{\mu(T^{-1}A_j)} = \frac{\mu(T^{-1}A_k \cap A_j)}{\mu(A_j)}$$

by T-invariance of μ, so we need only show that

$$\frac{\mu(T^{-2}A_k \cap T^{-1}A_j \cap A_i)}{\mu(T^{-1}A_j \cap A_i)} = \frac{\mu(T^{-1}A_k \cap A_j)}{\mu(A_j)} \tag{32}$$

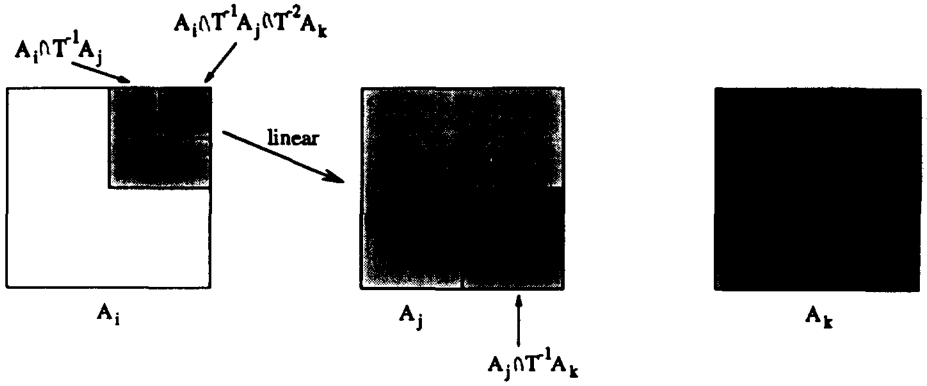

Fig. 2. The ratio of the measures of the heavily shaded regions to the measures of the shaded regions is preserved.

We note that $T(A_i \cap T^{-1}A_j) = A_j$ and $T(A_i \cap T^{-1}A_j \cap T^{-2}A_k) = A_j \cap T^{-1}A_k$ because $\{A_1, \ldots, A_n\}$ is a Markov partition. Since T is linear on each A_i, and the density of μ is piecewise constant on each A_i (Boyarsky and Haddard [20]), the ratio of the measures of the heavily shaded regions to the lightly shaded regions in Fig. 2 is preserved, and equation (32) holds. ■

Since $\{x_i : i = 0, 1, 2, \ldots\}$ is a Markov process, we have that

$$(P^k)_{ij} = \mathrm{Prob}\{x_k = j \mid x_0 = i\} = \mathrm{Prob}\{T^k x \in A_j \mid x \in A_i\}$$
$$= \frac{\mu(A_i \cap T^{-k}A_j)}{\mu(A_i)}. \tag{33}$$

We may, however, show (33) directly by induction.

Assume that $(P^k)_{ij} = \mu(A_i \cap T^{-k}A_j)/\mu(A_i)$. Then

$$\begin{aligned}
(P^{k+1})_{il} = \sum_{j=1}^{n} (P^k)_{ij} P_{jl} &= \sum_{j=1}^{n} \frac{\mu(A_i \cap T^{-k}A_j)}{\mu(A_i)} \frac{\mu(A_j \cap T^{-1}A_l)}{\mu(A_j)} \\
&= \sum_{j=1}^{n} \frac{\mu(A_i \cap T^{-k}A_j)}{\mu(A_i)} \frac{\mu(T^{-k}A_j \cap T^{-(k+1)}A_l \cap A_i)}{\mu(T^{-k}A_j \cap A_i)} \quad \text{by (31)} \\
&= \sum_{j=1}^{n} \frac{\mu(T^{-(k+1)}A_l \cap T^{-k}A_j \cap A_i)}{\mu(A_i)} \\
&= \frac{\mu(A_i \cap T^{-(k+1)}A_l)}{\mu(A_i)} \\
&= \mathrm{Prob}\{T^{k+1}x \in A_l \mid x \in A_i\},
\end{aligned}$$

and so by induction, we obtain (33).

As we refine our partition, we insist that it remains Markov, that is, it satisfies (i)-(iii) of Definition 7.1. To this end, we refine our partition $\mathfrak{B}$ by taking joins with successive inverse images of itself. Set the refined partition $\mathfrak{B}' = \mathfrak{B} \vee T^{-1} \vee \cdots \vee T^{-l}\mathfrak{B}$ (of cardinality n_l) and denote an element of $\mathfrak{B}'$ by $A_{i'}$, $i' \in \{1, \ldots, n_l\}$, where $A_{i'} = A_{i_1} \cap T^{-1}A_{i_2} \cap \cdots \cap T^{-l}A_{i_{l+1}}$, $i_k \in S$, $k = 1, \ldots, l+1$. Our transition matrix for this refined partition is defined by

$$P'_{i'j'} := \frac{\mu(A_{i'} \cap T^{-1}A_{j'})}{\mu(A_{i'})}.$$

We are interested in how much longer it takes our refined system to reach equilibrium as compared with the system corresponding to the original partition $\mathfrak{B}$.

LEMMA 7.3.

$$\max_{1 \le i' \le n_l} \|(P')^{k+l}(i', \cdot) - p'(\cdot)\|_1 = \max_{1 \le i \le n} \|P^k(i, \cdot) - p(\cdot)\|_1.$$

Proof.

$$\begin{aligned}
&((P')^{k+l})_{i'j'} \\
&\quad = \mathrm{Prob}\{T^{k+l}x \in A_{j'} \mid x \in A_{i'}\} \\
&\quad = \mathrm{Prob}\{T^{k+l}x \in A_{j_1}, \ldots, T^{-k+2l}x \in A_{j_{l+1}} \mid x \in A_{i_1}, Tx \in A_{i_2}, \ldots, T^l x \in A_{i_{l+1}}\} \\
&\quad = \mathrm{Prob}\{x_{k+2l} = j_{l+1}, \ldots, x_{k+l+1} = j_2, x_{k+l} = j_1 \mid x_l = i_{l+1}, \ldots, x_1 = i_2, x_0 = i_1\} \\
&\qquad\qquad \text{(in random variable notation)} \\
&\quad = \mathrm{Prob}\{x_{k+2l} = j_{l+1}, \ldots, x_{k+l+1} = j_2, x_{k+l} = j_1 \mid x_l = i_{l+1}\} \quad \text{by the Markov property,} \\
&\quad = \mathrm{Prob}\{x_{k+2l} = j_{l+1} \mid x_{k+2l-1} = j_l\} \cdots \mathrm{Prob}\{x_{k+l+1} = j_2 \mid x_{k+l} = j_1\} \\
&\qquad \cdot \mathrm{Prob}\{x_{k+l} = j_1 \mid x_l = i_{l+1}\} \\
&\quad = P_{j_l j_{l+1}} \cdots P_{j_1 j_2} (P^k)_{i_{l+1} j_1}.
\end{aligned}$$

Now we see that

$$\begin{aligned}
\max_{i'} \|(P')^{k+l}(i', \cdot) - p'(\cdot)\|_1 &= \max_{i'} \sum_{j'} |(P')^{k+l}(i', j') - p'(j')| \\
&= \max_{i_1, \ldots, i_{l+1}} \sum_{j_1, \ldots, j_{l+1}} |(P^k)_{i_{l+1} j_1} P_{j_1 j_2} \cdots P_{j_l j_{l+1}} - p_{j_1} P_{j_1 j_2} \cdots P_{j_l j_{l+1}}| \\
&= \max_{i_1, \ldots, i_{l+1}} \sum_{j_1, \ldots, j_{l+1}} P_{j_1 j_2} \cdots P_{j_l j_{l+1}} |(P^k)_{i_{l+1} j_1} - p_{j_1}| \\
&= \max_{i_{l+1}} \sum_{j_1} |(P^k)_{i_{l+1} j_1} - p_{j_1}| \\
&= \max_i \|P^k(i, \cdot) - p(\cdot)\|_1, \quad \blacksquare
\end{aligned}$$

Thus if our partition is refined by taking joins with l inverse iterates of the original partition, then after an extra l iterations of the "refined" random system, we are just as far from equilibrium as we would be in the original random system. This is roughly because we may make our starting distributions more concentrated by putting all the initial mass in a smaller refined partition set. It then takes longer for this more concentrated distribution to spread out.

As before, we find constants R, $r > 0$ such that

$$\max_{1 \le i \le n} \|P^k(i,\cdot) - p(\cdot)\|_1 \le Rr^k \qquad \text{for all } k \ge 0. \tag{34}$$

After refining $\mathfrak{B}$ to obtain $\mathfrak{B}'$, we expect the Markov system generated by the corresponding transition matrix P' to take longer to settle to equilibrium. We shall show that the exponential rate of convergence to equilibrium is the same for each refinement of $\mathfrak{B}$, and that the corresponding constants R_n grow polynomially with the cardinality of the partition n.

PROPOSITION 7.4. As we refine our Markov partitions R_n grows polynomially with n and $r_n = r_0$ for all $n \ge 0$.

Proof. Let R_0 be the constant associated with the original partition $\mathfrak{B}$ and R_l be the constant for the partition $\mathfrak{B}'$. We have that

$$\max_{1 \le i' \le n_l} \|(P')^{k+l}(i', \cdot) - p'(\cdot)\| = \max_{1 \le i \le n} \|P^k(i, \cdot) - p(\cdot)\| \le R_0 r_0^k = (R_0/r_0^l) r_0^{k+l},$$

for all $k, l \ge 0$. Thus we may take $R_l = R_0/r_0^l$ and $r_l = r_0$. This tells us how R_l grows with the number of inverse image refinements, but we really want to know how it grows with the number of partition sets.

Put $\lambda_{\min} = \inf_x \|D_x T\|$, noting that $1 < \lambda_{\min}$. There exists a constant C_1 such that $C_1 \lambda_{\min}^l \operatorname{card} \mathfrak{B} \le \operatorname{card} \mathfrak{B}'$, so $\lambda_{\min}^l \le C_1^{-1} n_l/n_0$, where $n_0 = \operatorname{card} \mathfrak{B}$ and $n_l = \operatorname{card} \mathfrak{B}'$. We choose $q > 0$ such that $1/r_0 < \lambda_{\min}^q$. Then

$$R_l = R_0 r_0^{-l} \le R_0 \lambda_{\min}^{ql} \le R_0 \left(\frac{n_l}{C_1 n_0}\right)^q = \left(\frac{R_0}{(C_1 n_0)^q}\right) n_l^q,$$

and R_n grows polynomially with n as required. ■

Proposition 6.8 cannot be directly applied to this system, as the sequence of Markov partitions $\mathfrak{B}$ does not satisfy Assumption 3.1(ii) and (iii). This is because the size of the Markov partition sets shrink at different geometric rates. Assumption (ii) may be suitably modified to read:

(ii′) There is a constant α such that $T(A_i^n)$ covers at most α of the sets $\{A_1^n, \ldots, A_n^n\}$ for each $i = 1, \ldots, n$,

without changing the proof of Lemma 3.9. In this situation, the partitions are tied to the dynamics of T, so that at each refinement the maximum number of sets an image of a partition set can cover is bounded by the same number. The algebraic lower bound of Assumption (iii) for $\mu(A_i^n)$ cannot be simply fixed; indeed $\min_i \mu(A_i^n)$ decreases geometrically with n as the partitions are refined. This does not really matter though, as our invariant density h is piecewise constant on each partition set, the difference $|\tilde{P}_{n,ij} - P_{n,ij}|$ in Lemma 3.6 is in fact zero, as is the first term of (29) in the proof of Proposition 6.8. Thus we obtain the exact answer μ at the first approximation stage. The fact that the sequence of invariant measures $\tilde{\mu}_n$ defined by (7) converge weakly to the unique absolutely continuous invariant measure μ follows as a special case of the main theorem of [7].

The situation described in this section is a rather trivial one, since μ may be computed *exactly* without using any of our extra perturbation machinery. This is because the Perron–Frobenius operator has an *exact* finite dimensional description. We have, however, learnt a good deal more about our matrix construction and have given an example of when we have complete understanding of how the constants R_n and r_n behave.

7.2. *Piecewise $C^{1+\mathrm{Lip}}$ expanding maps of [0, 1]*

In this section we describe another setting where we have control of the constants R_n and r_n. $T: [0, 1] \circlearrowleft$ is piecewise $C^{1+\mathrm{Lip}}$ expanding and possesses a unique invariant measure μ equivalent to Lebesgue with density h. We restrict the Perron–Frobenius operator to the Banach space of functions of bounded variation $(BV, \|\cdot\|)$ where $BV = \{f \in L^1([0, 1], m) : \mathrm{var} f < \infty\}$ and $\|\cdot\| = \|\cdot\|_1 + \mathrm{var}(\cdot)$. Lasota and Yorke [21] guarantee the existence of an absolutely continuous invariant measure μ; we assume that it is unique and equivalent to Lebesgue (by making T mixing with respect to μ, for example Keller [22]). By the uniqueness assumption on μ, $\mathcal{P}$ has exactly one fixed point h.

We outline what we are about to do. Our maps T will have a large and slowly varying derivative in order that the Perron–Frobenius operator is a contraction in the $\|\cdot\|$ norm (restricted to test functions with zero mean). Our finite rank approximation will be a "projection" of the Perron–Frobenius operator acting on some finite dimensional function space. Restricted to this finite dimensional space, $\mathcal{P}$ may be represented by our matrix P defined in (8). The norm of our projection is bounded above by unity, so that our finite rank operator is also a contraction in the $\|\cdot\|$ norm when restricted to test functions of zero integral. From this, we immediately obtain a bound for our "R_n's" and "r_n's" for our matrix approximation.

7.2.1. Preliminary lemmas. Some of the initial lemmas are based on ideas in Rychlik [23]. It will become apparent later that we must deal with $\mathcal{P}_\mu$, the Perron–Frobenius operator with respect to the unique absolutely continuous invariant measure μ, defined by

$$\mathcal{P}_\mu f(x) = \sum_{z \in T^{-1}x} \frac{f(z)h(z)}{|\det DT(z)|h(Tz)}. \tag{35}$$

This operator may be also defined by

$$\int_A \mathcal{P}_\mu f \, d\mu = \int_{T^{-1}A} f \, d\mu, \qquad \text{for all measurable } A \subset [0, 1]. \tag{36}$$

From (36), it is easy to see that $f \equiv 1$ is an eigenfunction of eigenvalue one. Compare equations (35) and (36) with the more usual definition of the Perron–Frobenius operator with respect to Lebesgue measure in (4) and (5). Note that $\|f\|_1$ is given by $\int |f| \, d\mu$ in this subsection. In all of what follows, $g(x) = h(x)/(|T'(x)| \cdot h(Tx))$; in Rychlik's paper, g is simply $1/|T'|$. Denote by $\mathbb{C}_\mu^\perp$ those functions with zero μ-mean, that is,

$$\mathbb{C}_\mu^\perp = \left\{ f \in BV : \int f \, d\mu = 0 \right\}. \tag{37}$$

SUBLEMMA 7.5. If $f \in \mathbb{C}_\mu^\perp$, then $\operatorname{var} f \geq \|f\|_\infty \geq \|f\|_1$.

Proof. Assume $f \in \mathbb{C}_\mu^\perp$ is not identically equal to zero. There exists $x^+ \in [0, 1]$ such that the positive part of f, namely $f^+ = \max\{f, 0\}$, is zero at x^+. For if not, $f > 0$ for all $x \in [0, 1]$ and $\int f \, d\mu > 0$. Similarly, the graph of f^- touches the x-axis. Thus

$$\operatorname{var} f^+ \geq \sup f^+ \geq \int f^+ \, d\mu,$$

and

$$\operatorname{var} f^- \geq \sup f^- \geq \int f^- \, d\mu = \int f^+ \, d\mu.$$

Noticing that $\operatorname{var} f = \operatorname{var} f^+ + \operatorname{var} f^-$, one has

$$\operatorname{var} f \geq \sup f^+ + \sup f^- \geq \sup |f| \geq \int |f| \, d\mu,$$

and we are done. ■

Next we give a bound for the norm of $\mathcal{P}_\mu$, for expanding maps of the circle.

LEMMA 7.6. Let $T: S^1 \circlearrowleft$ be C^1 and expanding. Then for $f \in \mathbb{C}_\mu^\perp$,

$$\operatorname{var} \mathcal{P}_\mu f \leq (\|g\|_\infty + \operatorname{var} g) \operatorname{var} f \tag{38}$$

and

$$\|\mathcal{P}_\mu|_{\mathbb{C}_\mu^\perp}\| \leq 2(\|g\|_\infty + \operatorname{var} g). \tag{39}$$

Proof. If we plot the graph of T on [0, 1], identifying the endpoints, the graph consists of a finite number of branches, progressing from 0 at the bottom to 1 at the top, with the next branch reappearing at 0 directly below where the previous one was 1. These branches partition the unit interval into subintervals $\beta = \{B_1, \ldots, B_r\}$, with $T(B_i) = [0, 1]$, $i = 1, \ldots, r$.

Let $B \in \beta$. To begin with, we note that the action of $\mathcal{P}_\mu$ on $f \cdot \chi_B$ is roughly to "multiply it by g and stretch it across the whole unit interval". This is just the push forward of the density $f \cdot \chi_B$. Precisley, $\mathcal{P}_\mu(f \cdot \chi_B)(x) = f(T^{-1}x) \cdot g(T^{-1}x)$, where the single branch of T^{-1} corresponding to B is taken. As $T|_B$ is monotonic, clearly,

$$\operatorname{var}(f \circ T|_B^{-1} \cdot g \circ T|_B^{-1}) = \operatorname{var}_{TB}(f \circ T|_B^{-1} \cdot g \circ T|_B^{-1}) = \operatorname{var}_B f \cdot g, \tag{40}$$

noting that $TB = [0, 1]$; see [1] for properties of variation. Let $f \in \mathbb{C}_\mu^\perp$,

$$\begin{aligned}
\operatorname{var} \mathcal{P}_\mu f &= \operatorname{var} \sum_{B \in \beta} \mathcal{P}_\mu(f \cdot \chi_B) && \text{by linearity of } \mathcal{P}_\mu \\
&\leq \sum_{B \in \beta} \operatorname{var} \mathcal{P}_\mu(f \cdot \chi_B) && \text{as } \operatorname{var}(f_1 + f_2) \leq \operatorname{var} f_1 + \operatorname{var} f_2 \\
&= \sum_{B \in \beta} \operatorname{var}_B(f \cdot g) && \text{by (40)} \\
&= \operatorname{var}(f \cdot g) \\
&\leq \operatorname{var} f \|g\|_\infty + \|f\|_\infty \operatorname{var} g && \text{general inequality} \\
&= (\|g\|_\infty + \operatorname{var} g) \operatorname{var} f && \text{by Sublemma 7.5.} \qquad (41)
\end{aligned}$$

Also

$$\begin{aligned}
\|\mathcal{P}_\mu|_{\mathbb{C}_\mu^\perp}\| &= \sup_{f \in \mathbb{C}_\mu^\perp} \frac{\|\mathcal{P}_\mu f\|_1 + \operatorname{var} \mathcal{P}_\mu f}{\|f\|_1 + \operatorname{var} f} \\
&\leq \sup_{f \in \mathbb{C}_\mu^\perp} \frac{2 \operatorname{var} \mathcal{P}_\mu f}{\operatorname{var} f} && \text{as } \mathcal{P}_\mu f \in \mathbb{C}_\mu^\perp, \text{ and by Sublemma 7.5} \\
&\leq 2(\|g\|_\infty + \operatorname{var} g). \quad \blacksquare
\end{aligned}$$

The next step is to talk about piecewise $C^{1+\mathrm{Lip}}$ expanding maps of [0, 1] whose branches do not all progress from the bottom to the top of the graph. The main difference in the analysis of this case is that equality in (41) becomes an inequality (the wrong way round), as we have to worry about the jumps at the endpoints of TB.

LEMMA 7.7. Let $T: [0, 1] \circlearrowleft$ be piecewise $C^{1+\mathrm{Lip}}$ and expanding. Denote by N_1 the number of branches that start at either the top or bottom of the graph of T but do not traverse the unit interval fully, and by N_2 the number of branches that meet neither the top nor the bottom of the graph, see Fig. 3. Then for $f \in \mathbb{C}_\mu^\perp$,

$$\operatorname{var} \mathcal{P}_\mu f \leq ((2N_2 + N_1 + 1)\|g\|_\infty + \operatorname{var} g) \operatorname{var} f, \tag{42}$$

and

$$\|\mathcal{P}_\mu|_{\mathbb{C}_\mu^\perp}\| \leq 2((2N_2 + N_1 + 1)\|g\|_\infty + \operatorname{var} g) \tag{43}$$

Proof. As in the proof of Lemma 7.6, we have $\operatorname{var} \mathcal{P}_\mu f \leq \sum_{B \in \beta} \operatorname{var} \mathcal{P}_\mu(f \cdot \chi_B)$. Recall that $\mathcal{P}_\mu(f \cdot \chi_B)$ was "$f \cdot \chi_B$ stretched out and multiplied by g". In the proof of Lemma 7.6, TB was the entire unit interval, so that $\operatorname{var} \mathcal{P}_\mu(f \cdot \chi_B)$ was equal to $\operatorname{var}_B(f \cdot g)$. However, in our more general situation, we must add in the jumps at the endpoints of TB. So our new inequality is

$$\operatorname{var} \mathcal{P}_\mu(f \cdot \chi_B) \leq \operatorname{var}_B(f \cdot g) + \Theta \|f\|_\infty \|g\|_\infty,$$

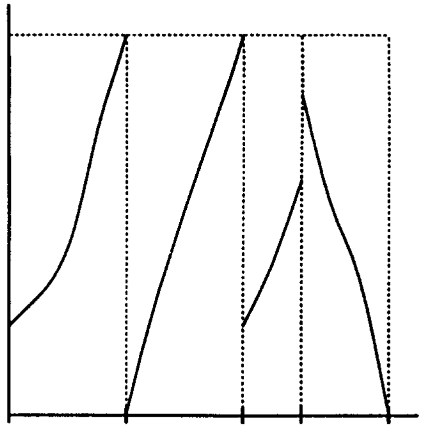

Fig. 3. Graph of an interval map T. In this example, $N_0 = 1$, $N_1 = 2$ and $N_3 = 1$.

where $\Theta = 1$ if TB has one of 0 or 1 as an endpoint and $\Theta = 2$ if the endpoints of TB are neither 0 nor 1. If one endpoint of TB is 0 and the other is 1, then $\Theta = 0$ and we are in the situation of Lemma 7.6. Thus

$$\begin{aligned}\operatorname{var} \mathcal{P}_\mu f &\le \sum_{B\in\beta} \operatorname{var} \mathcal{P}_\mu(f\cdot\chi_B)\\ &\le \sum_{B\in\beta} \operatorname{var}_B(f\cdot g) + (2N_2+N_1)\|f\|_\infty\|g\|_\infty\\ &\le (\|g\|_\infty + \operatorname{var} g)\operatorname{var} f + (2N_2+N_1)\|g\|_\infty \operatorname{var} f\\ &= ((2N_2+N_1+1)\|g\|_\infty + \operatorname{var} g)\operatorname{var} f.\end{aligned}$$

Inequality (43) follows as in Lemma 7.6. ■

Definition 7.8. Define the averaging operator $\mathcal{A}_\mu : BV \to \text{Const}$, where Const is the one-dimensional space of real-valued constant functions on $[0, 1]$, by

$$(\mathcal{A}_\mu f)(x) = \left(\int f\,\mathrm{d}\mu\right)\chi_{[0,1]}(x).$$

LEMMA 7.9. Let T be as in Lemma 7.7 and suppose that

$$\gamma := 2((2N_2+N_1+1)\|g\|_\infty + \operatorname{var} g) < 1.$$

Then the operator $\mathcal{P}_\mu - \mathcal{A}_\mu$ is a contraction on $(BV, \|\cdot\|)$ of factor γ.

Proof. Let $f \in BV$ be given, and note that $(\mathcal{P}_\mu - \mathcal{A}_\mu)f = \mathcal{P}_\mu f - \int f\,\mathrm{d}\mu = \mathcal{P}_\mu(f - \int f\,\mathrm{d}\mu)$ as $\mathcal{P}_\mu 1 = 1$. Put $f_0 = f - \int f\,\mathrm{d}\mu$, and apply Lemma 7.7 to f_0. One obtains $\|(\mathcal{P}_\mu - \mathcal{A}_\mu)f\| = \|\mathcal{P}_\mu f_0\| < \gamma\|f_0\| \le \gamma\|f\|$, with the final inequality following as $\operatorname{var} f_0 = \operatorname{var} f$ and $\|f_0\|_1 \le \|f\|_1$. ■

7.2.2. Bounding the norm of the matrix approximation. We suppose that our map $T: [0, 1] \circlearrowleft$ is expanding so strongly and uniformly, that $\mathcal{P}_\mu$ restricted to functions with zero μ-mean is an immediate contraction in the $\|\cdot\| = \|\cdot\|_1 + \mathrm{var}(\cdot)$ norm. In other words, $\|\mathcal{P}_\mu|_{C_\mu^\perp}\| \le \gamma < 1$. Clearly $\|\mathcal{P}_\mu - \mathcal{Q}_\mu\| \le \gamma$ also.

Definition 7.10. Given some partition of $[0, 1]$ into subintervals $\{A_1, \ldots, A_n\}$, we define

$$F = \left\{ f \in BV : f = \sum_{k=1}^{n} a_k \chi_{A_k},\ a_k \in \mathbb{R},\ k = 1, \ldots, n \right\},$$

a finite dimensional vector space of piecewise constant functions, constant on each A_k. Further define a projection $\pi_\mu : BV \to F$ by

$$\pi_\mu(f) = \sum_{k=1}^{n} \left(\frac{1}{\mu(A_k)} \int_{A_k} f \,\mathrm{d}\mu \right) \chi_{A_k},$$

where $\pi_\mu(f)$ on A_k takes on the value of f averaged over A_k.

Note that $\mathcal{P}_\mu \mathcal{Q}_\mu = \mathcal{Q}_\mu$, $\pi_\mu \mathcal{Q}_\mu = \mathcal{Q}_\mu$, that $\pi_\mu \mathcal{P}_\mu$ is a finite-rank operator, and that $\|\pi_\mu\| \le 1$. The latter is clear as $\mathrm{var}\, \pi_\mu f \le \mathrm{var} f$, and $\|\pi_\mu f\|_1 \le \|f\|_1$. Consider $\pi_\mu(\mathcal{P}_\mu - \mathcal{Q}_\mu)$ and note that $\|\pi_\mu(\mathcal{P}_\mu - \mathcal{Q}_\mu)\| \le \|\pi_\mu\| \|\mathcal{P}_\mu - \mathcal{Q}_\mu\| \le 1 \cdot \gamma$. We wish to restrict $\pi_\mu \mathcal{P}_\mu$ and $\pi_\mu \mathcal{Q}_\mu$ to the space F so that we may obtain a matrix representation for them with respect to the basis of F, namely the characteristic functions $\{\chi_{A_1}, \ldots, \chi_{A_n}\}$. We will denote the matrix representations by $[\pi_\mu \mathcal{P}_\mu]$ and $[\mathcal{Q}_\mu]$, respectively.

LEMMA 7.11. Under left multiplication, $[\pi_\mu \mathcal{P}_\mu]$ and $[\mathcal{Q}_\mu]$ have matrix representation†

$$[\pi_\mu \mathcal{P}_\mu]_{ij} = \frac{\mu(A_i \cap T^{-1} A_j)}{\mu(A_j)}, \tag{44}$$

and

$$[\mathcal{Q}_\mu]_{ij} = \mu(A_i). \tag{45}$$

Proof. Let $f = \sum_{i=1}^{n} a_i \chi_{A_i} \in F$.

$$\begin{aligned}
\pi_\mu \mathcal{P}_\mu \left(\sum_{i=1}^{n} a_i \chi_{A_i}(x) \right) &= \sum_{j=1}^{n} \left(\frac{1}{\mu(A_j)} \int_{A_j} \left(\mathcal{P}_\mu \left(\sum_{i=1}^{n} a_i \chi_{A_i}(x) \right) \right) \mathrm{d}\mu \right) \chi_{A_j} \\
&= \sum_{j=1}^{n} \left(\frac{1}{\mu(A_j)} \int_{T^{-1}A_j} \sum_{i=1}^{n} a_i \chi_{A_i} \,\mathrm{d}\mu \right) \chi_{A_i} \\
&= \sum_{j=1}^{n} \underbrace{\left(\sum_{i=1}^{n} a_i \frac{\mu(A_i \cap T^{-1} A_j)}{\mu(A_j)} \right)}_{a_j'} \chi_{A_j}.
\end{aligned}$$

So $a_j' = \sum_i a_i [\pi_\mu \mathcal{P}_\mu]_{ij}$, where $a_j' := [(\pi_\mu \mathcal{P}_\mu) F]_j$.

† This has been noticed by Li [12] for the standard Perron-Frobenius operator with respect to *Lebesgue* measure.

To obtain the matrix representation for $\mathcal{Q}_\mu$, we proceed as follows. As before, let $f = \sum_{i=1}^n a_i \chi_{A_i} \in F$.

$$\begin{aligned}\mathcal{Q}_\mu\left(\sum_{k=1}^n a_k \chi_{A_k}\right) &= \left(\int \sum_{k=1}^n a_k \chi_{A_k}\, \mathrm{d}\mu\right)\chi_{[0,1]}\\ &= \left(\sum_{k=1}^n \int_{A_k} a_k\, \mathrm{d}\mu\right)\chi_{[0,1]}\\ &= \left(\sum_{k=1}^n a_k \mu(A_k)\right)\chi_{[0,1]}.\end{aligned}$$

Thus $[\mathcal{Q}_\mu]_{ij} = \mu(A_i)$, under left multiplication. ■

Our goal is to describe how the R_n and r_n in equation (17) vary. To do this we need to transfer the information about the difference $(\pi_\mu \mathcal{P}_\mu - \mathcal{Q}_\mu)^k = (\pi_\mu \mathcal{P}_\mu)^k - \mathcal{Q}_\mu$† in the BV norm to information about the difference $[\pi_\mu \mathcal{P}_\mu]^k - [\mathcal{Q}_\mu]$ in an appropriate matrix norm (the standard L_1 matrix norm, for example).

We may represent an element of F in vector form as an ordered n-tuple of the step function values $a_1, \ldots, a_n$. To begin with, we define a vector norm $\|\cdot\|_w$ by $\|[a_1, \ldots, a_n]\| = \|f\|_1$, where $f = \sum_{i=1}^n a_i \chi_{A_i}$ (keeping in mind that $\|\cdot\|_1$ is integration of $|f|$ with respect to μ). This vector norm will induce a matrix norm in the standard way, which we also denote by $\|\cdot\|_w$. Now if a step function $f \in F$ has $\|\cdot\|_w$ norm 1, how big can its $\|\cdot\|$ norm be? Well, by definition, $\|f\|_1 = 1$, and one sees that $\operatorname{var} f \le 2\|f\|_1/\min_{1\le i\le n} \mu(A_i)$. Thus for $f \in F$, $\|f\| \le \|f\|_w + 2\|f\|_w/\min_{1\le i\le n}\mu(A_i)$, and so

$$\left(1 + 2/\min_{1\le i\le n} \mu(A_i)\right)^{-1}\|f\| \le \|f\|_w \le \|f\|.$$

Let $f \in F$, and in what follows, continue to think of f as a step function and a row vector interchangeably. When we write $[\pi_\mu \mathcal{P}_\mu] f$ we mean left multipication of the matrix $[\pi_\mu \mathcal{P}_\mu]$ by the row vector $[a_1, \ldots, a_n]$ where $f = \sum_i a_i \chi_{A_i}$.

$$\begin{aligned}\|[\pi_\mu \mathcal{P}_\mu]^k - [\mathcal{Q}_\mu]\|_w &= \sup_{f\in F} \frac{\|([\pi_\mu \mathcal{P}_\mu]^k - [\mathcal{Q}_\mu])f\|_w}{\|f\|_w}\\ &= \sup_{f\in F} \frac{\|([\pi_\mu \mathcal{P}_\mu] - [\mathcal{Q}_\mu])^k f\|_w}{\|f\|_w} \qquad \text{as in the proof of Theorem 6.4}\\ &\le \sup_{f\in F} \frac{\|(\pi_\mu \mathcal{P}_\mu - \mathcal{Q}_\mu)^k f\|}{(1 + 2/\min_{1\le i\le n}\mu(A_i))^{-1}\|f\|}\\ &\le \sup_{f\in F} \frac{(1 + 2/\min_{1\le i\le n}\mu(A_i))\|(\pi_\mu \mathcal{P}_\mu - \mathcal{Q}_\mu)\|^k\|f\|}{\|f\|}\\ &\le \left(1 + 2/\min_{1\le i\le n}\mu(A_i)\right)\gamma^k. \end{aligned} \tag{46}$$

The observant reader may have noticed that the matrix $[\pi_\mu \mathcal{P}_\mu]_{ij}$ is not exactly the matrix P_{ij} we were studying in earlier sections. We must perform a similarity transformation on

† To show this equality, one follows the argument for matrices given in the proof of Theorem 6.4 in [16].

$[\pi_\mu \mathcal{P}_\mu]$ and $[\mathcal{A}_\mu]$ to make these matrices stochastic. We will use a different norm $\|\cdot\|_m$ for these matrices, this time the matrix norm induced by the standard L^1 vector norm, namely $\|[a_1, \ldots, a_n]\|_m = \sum_{i=1}^n |a_i|$. In fact for a matrix B, $\|B\|_m = \|B\|_1$ in our earlier notation, but we use $\|\cdot\|_m$ to avoid confusion with the $\|\cdot\|_1$ norm for functions. Let us suppose that our $f \in F$ has $\|\cdot\|_w$ norm 1 and ask "how big can $\|f\|_m$ be?". It is easy to see that the most $\|f\|_m$ can be is $1/\min_{1\le i\le n} \mu(A_i)$ (we put all the mass of f into the smallest partition set to get the greatest height). Thus $\|f\|_m \le 1/\min_{1\le i\le n} \mu(A_i)\|f\|_w$. Now how small can $\|f\|_m$ be? By similar reasoning, $\|f\|_m \ge 1/\max_{1\le i\le n} \mu(A_i)\|f\|_w$, so putting $c_n = \min_{1\le i\le n} \mu(A_i)$ and $C_n = \max_{1\le i\le n} \mu(A_i)$ we have

$$C_n^{-1}\|f\|_w \le \|f\|_m \le c_n^{-1}\|f\|_w.$$

We define the diagonal matrix $Q_{ij} = \mu(A_i)\delta_{ij}$, and the stochastic matrix

$$P_{ij} := (Q^{-1}[\pi_\mu \mathcal{P}_\mu]Q)_{ij} = \frac{\mu(A_i \cap T^{-1}A_j)}{\mu(A_i)}.$$

We also define

$$A_{ij} := (Q^{-1}[\mathcal{A}_\mu]Q)_{ij} = \mu(A_j).$$

Note that $A_{ij} = p_j$ where p is the fixed left eigenvector of P. We wish to bound $\|P^k - A\|_m$. Note first that $\|Q\|_m = \max_{1\le i\le n} \mu(A_i)$ and $\|Q^{-1}\|_m = 1/\min_{1\le i\le n} \mu(A_i)$ so that

$$\|Q\|_w = \sup_{f\in F} \frac{\|Qf\|_w}{\|f\|_w} \le \sup_{f\in F} \frac{C_n\|Qf\|_m}{c_n\|f\|_m} \le \beta \max_{1\le i\le n} \mu(A_i),$$

where we have used Assumption 3.1(iii). Reverting to our old L^1 matrix norm notation, we have

$$\max_{1\le i\le n} \|P^k(i, \cdot) - p(\cdot)\|_1 = \|P^k - A\|_m.$$

Now,

$$\begin{aligned}
\|P^k - A\|_m &= \sup_{f\in F} \frac{\|(P^k - A)f\|_m}{\|f\|_m} \\
&= \sup_{f\in F} \frac{\|Q^{-1}([\pi_\mu \mathcal{P}_\mu]^k - [\mathcal{A}_\mu])Qf\|_m}{\|f\|_m} \\
&\le \sup_{f\in F} \frac{\|Q^{-1}\|_m\|([\pi_\mu \mathcal{P}_\mu]^k - [\mathcal{A}_\mu])Qf\|_m}{\|f\|_m} \\
&\le \sup_{f\in F} \frac{c_n^{-1} \cdot c_n^{-1}\|[\pi_\mu \mathcal{P}_\mu]^k - [\mathcal{A}_\mu]Qf\|_w}{C_n^{-1}\|f\|_w} \\
&\le \sup_{f\in F} \frac{c_n^{-2}\|[\pi_\mu \mathcal{P}_\mu]^k - [\mathcal{A}_\mu]\|_w\|Q\|_w\|f\|_w}{C_n^{-1}\|f\|_w} \\
&\le c_n^{-2}(1 + 2c_n^{-1})\gamma^k \cdot C_n^2 \qquad \text{by (46)} \\
&\le \beta^2(1 + 2c_n^{-1})\gamma^k \\
&\le \beta^2(1 + 2\beta n)\gamma^k \qquad \text{as } 1/n \le C_n \le \beta c_n, \\
&= R_n r^k \qquad \text{where } R_n = \beta^2(1 + 2\beta n) \text{ and } r = \gamma.
\end{aligned}$$

Thus R_n has polynomial (in fact, linear) growth rate with n, as required. We now apply Proposition 6.8 to prove the following.

PROPOSITION 7.12. Suppose that T: $[0, 1]\circlearrowleft$ is a piecewise $C^{1+\mathrm{Lip}}$ expanding interval map such that $\gamma := 2((2N_2 + N_1 + 1)\|g\|_\infty + \operatorname{var} g) < 1$. Let $\{\tilde{P}_n\}_{n=n_0}^{\infty}$ be a sequence of $(n \times n)$ transition matrices generated by (6) from a sequence of interval partitions $\{\mathfrak{B}_n\}_{n=n_0}^{\infty}$ satisfying Assumption 3.1 whose maximal element diameters goes to zero as $n \to \infty$. The sequence of invariant measures $\{\tilde{\mu}_n\}_{n=n_0}^{\infty}$ defined by (7) converges strongly to the unique absolutely continuous μ of T. The rate of convergence is $O(\log n/n)$.

7.2.3. Extending the class of maps. The condition that $\gamma < 1$ in Lemma 7.9 is rather restrictive, in this section we briefly outline how to extend the class of maps to which our results may be applied. In order to decrease $\|g\|_\infty$ and $\operatorname{var} g$, we might consider T^k in the place of T for some $k > 1$. Setting $g_k(x) = h(x)/(|(T^k)'(x)|h(T^k x))$, we may make $\|g_k\|_\infty$ as small as we like by taking k large enough as T is expanding. However, there is no guarantee that $\operatorname{var} g_k$ will decrease as k becomes large, so we must abandon the simple bounds of Lemmas 7.6 and 7.7 and turn to some known results in spectral theory.

In our setting of piecewise $C^{1+\mathrm{Lip}}$ expanding mappings of the interval, the spectrum of the Perron–Frobenius operator $\mathcal{P}: BV\circlearrowleft$ consists of a finite number of eigenvalues on the unit circle, with the remainder bounded away from the unit circle; see Hofbauer and Keller [24]. As we are assuming that T is mixing, $\mathcal{P}$ will have just one eigenvalue on the unit circle of unit multiplicity, namely unity; see Keller [25]. We have an invariant splitting $BV = \{h\} \oplus \mathbb{C}_m^\perp$, where $\mathcal{P}h = h$ and $\mathbb{C}_m^\perp = \{f \in BV: \int f\,dm = 0\}$, so that $\mathcal{P}|_{\mathbb{C}_m^\perp}$ has spectral radius strictly less than one. Thus there exist constants $H > 0$ and $0 < q < 1$ such that $\|\mathcal{P}|_{\mathbb{C}_m^\perp}^k\| \leq Hq^k$. The relationship between the operators $\mathcal{P}$ and $\mathcal{P}_\mu$ is

$$\mathcal{P}_\mu^k f = \frac{\mathcal{P}^k(f \cdot h)}{h}, \qquad \text{for all } f \in L^1. \tag{47}$$

If $f \in \mathbb{C}_\mu^\perp$, then $f \cdot h \in \mathbb{C}_m^\perp$, and we choose k so large that $\|\mathcal{P}_\mu^k f\| = \|\mathcal{P}^k(f \cdot h)/h\| \leq H'q^k\|f\| = \gamma\|f\|$ for all $f \in \mathbb{C}_\mu^\perp$ and some $\gamma < 1$.

The idea is to apply Proposition 7.12 to T^k to approximate an absolutely continuous T^k-invariant measure μ^k. We have assumed that μ is the unique ACIM for T, and clearly μ is also T^k-invariant. Since T is mixing with respect to μ, T^k is also mixing and hence ergodic for $k \geq 1$. Thus, there can only be one invariant measure for T^k that is equivalent to Lebesgue, and it is $\mu = \mu^k$. All that we required to show the polynomial growth of the R_n's was that $\|\mathcal{P}_\mu|_{\mathbb{C}_\mu^\perp}\| < 1$, so we may now apply 7.12 to T^k and $\mathcal{P}_\mu^k$. We construct a matrix

$$\tilde{P}_{n,k,ij} = \frac{\mu(A_i \cap f^{-k}A_j)}{\mu(A_i)},$$

and compute $\tilde{p}_{n,k}$, the unique left eigenvector of unit eigenvalue. From $\tilde{p}_{n,k}$ we define a probability measure $\tilde{\mu}_n^k$ as in (7). Proposition 7.12 now states that $\tilde{\mu}_n^k \to \mu^k$ as $n \to \infty$ with the error being $O(\log n/n)$. As we have just noted, $\mu^k = \mu$. Similar ideas are contained in Hunt and Miller [26] who consider finite-dimensional approximations to quasi-compact†

† The Perron–Frobenius operator is called *quasi-compact* if it may be written as $\mathcal{P} = C + D$, where C has finite-dimensional range, simple eigenvalues on the unit circle in $\mathbb{C}$, and no other eigenvalues (except possibly 0), and the operator D has spectral radius strictly less than unity.

Perron–Frobenius operators of interval maps. The results in this subsection provide an alternative proof to the main result of [26]; one which avoids the use of sophisticated spectral perturbation theory.

7.3. *Numerical results for a higher dimensional Anosov diffeomorphism*

For general multidimensional Anosov diffeomorphisms, to obtain estimates or even bounds on the rate of mixing is very difficult. Further work in this direction is contained in the following chapter. In this section, we present some numerical results for a common linear automorphism of the 2-torus. Throughout, our map $T\colon \mathbb{T}^2 \circlearrowleft$ is defined by $T(x_1, x_2) = (2x_1 + x_2, x_1 + x_2) \pmod 1$. We initially partition the torus into 5 polygons, then refine this partition to 25, 50, 100, 200, 400 and 800 triangles. The partition of cardinality 50 is shown in Fig. 4 below, where the torus has been unwrapped onto the unit square, identifying opposing edges. To analyse this example, we will:

(i) directly evaluate the matrix 1-norm of the fundamental matrices Z_n for $n = 5$, 25, 50, 100, 200, 400 and 800 to demonstrate its logarithmic growth with n,

(ii) directly evaluate $\max_{1 \le i \le n} \|P_n^k(i, \cdot) - p_n(\cdot)\|_1$ for $n = 5$, 25, 50, 100, 200, 400 and 800 and $k = 1, \ldots, 30$ (see Fig. 6) to calculate:

(a) r_n and R_n, where $\max_{1 \le i \le n}\|P_n^k(i, \cdot) - p_n(\cdot)\|_1 \le R_n r_n^k$, to demonstrate that r_n is bounded above by some constant r', and that R_n grows polynomially with n.

(b) m_n, where m_n is the minimal m such that $\max_{1 \le i \le n} \|P_n^m(i, \cdot) - p_n(\cdot)\|_1 \le \min_i p_n(i)/2$, to demonstrate the logarithmic growth of m_n with n.

A direct evaluation of $\|Z_n\|_1$ is plotted against $\log_{10} n$ in the first plot of Fig. 5, for $n = 5$, 25, 50, 100, 200, 400 and 800. The presence of an almost linear graph suggests that in this case, $\|Z_n\|$ is of $O(\log n)$.

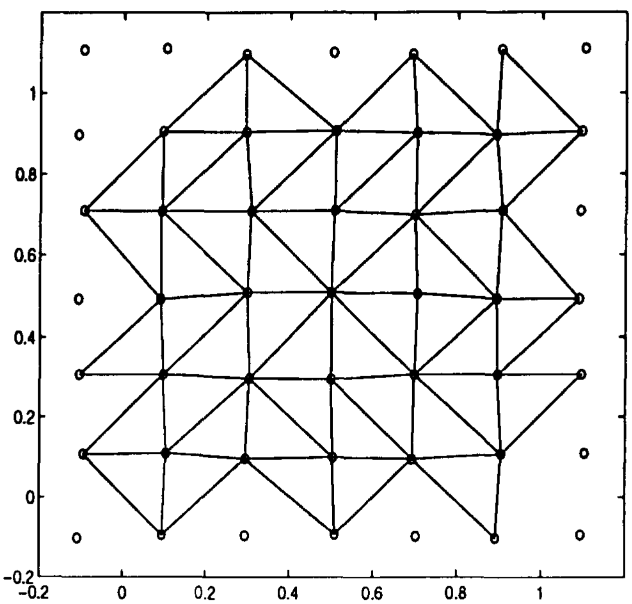

Fig. 4. The partition of cardinality 50 for the 2-torus. The torus has been unwrapped onto the unit square.

Table 1. Numerical tabulation of mixing indicators for Anosov's cat map

Number of states or triangles n	Matrix 1-norm of the fundamental matrix of P_n $\|Z_n\|_1$	y-intercept of the bounding lines R_n	Slope of the bounding lines r_n	Steps required to be "hal-fway" to equilibrium m_n
5	1.70	2.34	0.30	3
25	3.79	6.38	0.42	7
50	4.94	9.46	0.45	9
100	5.73	11.59	0.51	11
200	6.87	25.95	0.50	13
400	8.17	23.83	0.55	16
800	9.51	26.30	0.58	19

To calculate the minimal R_n and r_n, we first set r_n to equal the second largest (in magnitude) eigenvalue of P_n. This is because the second largest eigenvalue determines the rate of mixing for the Markov chain governed by P_n. We then choose R_n as small as possible, while still keeping $\max_i \|P_n^k(i, \cdot) - p_n(\cdot)\|_1 \leq R_n r_n^k$ for $k = 1, \ldots, 30$. The results of these calculations are shown in Table 1 and the second plot of Fig. 5. Notice that r_{200} is unusually small, and this is the reason that R_{200} has been forced up. This "glitch" is

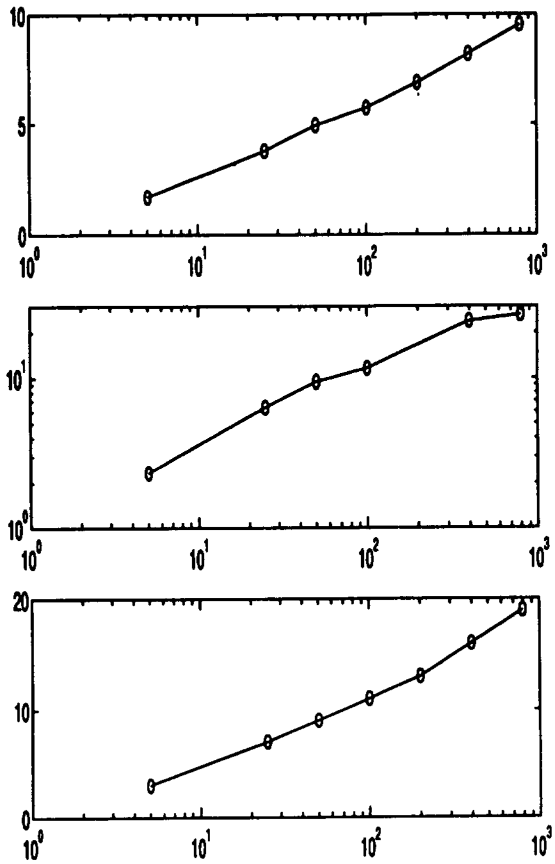

Fig. 5. Plots of $\|Z_n\|_1$, $\|R_n\|_1$ and m_n, respectively, vs n.

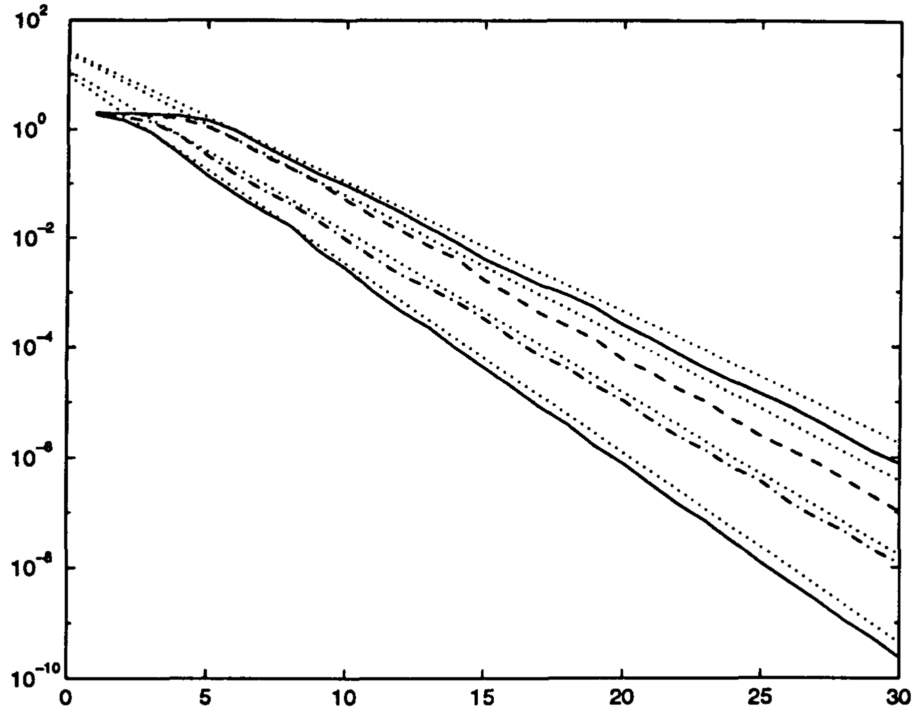

Fig. 6. Plot of $\max_{1 \le i \le n} \|P_n^m(i, \cdot) - p(\cdot)\|_1$ vs k for $n = 50$ (the lowest line), 100, 400, 800 (the highest line). The dotted lines represent minimal bounding lines of the form $R_n r_n^k$. The slope r_n is set to be the second largest (in magnitude) eigenvalue of P_n. Once r_n is fixed, R_n is chosen to be the minimal constant satisfying $\max_{1 \le i \le n} \|P_n^k(i, \cdot) - p(\cdot)\|_1 \le R_n r_n^k$ for $k = 1, \ldots, 30$.

an example of the complications involved with trying to draw relationships between the rate of mixing of a deterministic system and the rate of mixing of stochastic models of that deterministic system.

Calculating m_n was just a simple matter of increasing k until $\max_i \|P_n^k(i, \cdot) - p_n(\cdot)\|_1 \le \min_i p_n(i)/2$. These results are displayed in Table 1 and the third plot in Fig. 5. From the graphs it appears that R_n is growing polynomially and m_n is growing logarithmically with n as required.

We remark that even a linear toral automorphism with an arbitrary partition provides us with quite a complex system when it comes to estimating its rate of mixing. A linear system has been chosen so that Lebesgue measure is the absolutely continuous measure; this simplifies computations, but certainly does not mean that the results here are a special case. As our partition is in no way tied to the dynamics of T, we expect no extra complications with nonlinear Anosov diffeomorphisms.

DISCUSSION

The main thrust of this work is to present a new method for proving the conjecture of Ulam that the matrix $\tilde{P}$ in (6) is a good approximation of the Perron–Frobenius operator $\mathcal{P}$ of the map T in the sense that an estimate of the fixed point of $\mathcal{P}$ may be obtained from the fixed vector of $\tilde{P}$. This conjecture was first shown to be true for piecewise C^2

expanding mappings of the interval by Li [12]. More recently, Ding and Zhou [5, 6] produced a similar convergence result for multidimensional piecewise C^2 expanding maps. They used a higher order method with piecewise affine functions forming the finite dimensional basis rather than our piecewise constant basis. They also needed mild conditions on how the boundaries of the partition sets meet. The proofs of [12] and [5, 6] both rely on bounded variation arguments, and so may only be used to attack expanding mappings which "stretch" or "smooth out" density functions at each iteration. I believe that such strong behaviour may not be required of the map, and that Ulam's conjecture, or some variant, may also be true for maps that "mix" sufficiently quickly with respect to their physical measure. More precisely, maps for which an initial distribution of mass approaches the unique absolutely continuous invariant measure exponentially fast; uniformly hyperbolic systems such as Anosov diffeomorphisms enjoy such a property.

Our method is based on a few simple observations on the structure of the matrix approximation, and an appeal to perturbation theory of finite state Markov chains. The idea is that although our matrix $\tilde{P}$ is slightly different to the special matrix P, this error does not matter because the invariant density p is sufficiently insensitive to perturbation of the elements of P; thus $\tilde{p}$ is close to the "correct" density p. This robustness comes from the Markov chain associated with P being exponentially mixing. The difficulty comes in putting universal bounds on the rates of mixing of the Markov chains arising from the increasingly larger matrices P as we redefine our partitions of M. In some special cases, we have been able to obtain bounds using properties of the map T, but in general, to obtain rates of mixing of stochastic models of deterministic systems from the original system is a very difficult problem. Our numerical results for a more complicated higher dimensional map are encouraging however.

The advantages of this method are that no special partitions are required, one is not restricted to one-dimensional systems, and that only a modest mixing assumption is required of the map. We hope that this method provides a new avenue of opportunity for extending Ulam-type convergence results to a larger class of maps.

Acknowledgements—The author thanks Lai-Sang Young for helpful comments regarding Section 7.1. This work was partially supported by a grant from the Australian Research Council.

REFERENCES

1. Lasota, A. and Mackey, M. C., Chaos, fractals, and noise. Stochastic aspects of dynamics. *Applied Mathematical Sciences*, Vol. 97, 2nd edn. Springer-Verlag, New York, 1994.
2. Ulam, S. M., *Problems in Modern Mathematics*. Interscience, 1964.
3. Khas'minskii, R. Z., Principle of averaging for parabolic and elliptic differential equations and for Markov processes with small diffusion. *Theorem of Probability and its Applications*, 1963, **8**(1), 1-21.
4. Froyland, G., Judd, K., Mees, A. I., Murao, K. and Watson, D., Constructing invariant measures from data. *International Journal of Bifurcation and Chaos*, 1995, **5**(4), 1181-1192.
5. Jiu Ding and Ai Hui Zhou, The projection method for computing multidimensional absolutely continuous invariant measures. *Journal of Statistical Physics*, 1994, **77**(3/4), 899-908.
6. Jiu Ding and Ai Hui Zhou, Piecewise linear Markov approximations of Frobenius-Perron operators associated with multi-dimensional transformations. *Nonlinear Analysis*, 1995, **25**(4), 399-408.
7. Froyland, G., Finite approximation of Sinai-Bowen-Ruelle measures of Anosov systems in two dimensions. *Random & Computational Dynamics*, 1995, **3**(4), 251-264.
8. Mañé, R., *Ergodic Theory and Differentiable Dynamics*. Springer-Verlag, Berlin, 1987.
9. Bowen, R., Equilibrium states and the ergodic theory of Anosov diffeomorphisms. *Lecture Notes in Mathematics*, Vol. 470. Springer-Verlag, Berlin, 1975.

10. Lay, S. R., *Convex Sets and their Applications.* Pure & Applied Mathematics. Wiley, New York, 1982.
11. Conway, J. B., A course in funcitonal analysis. *Graduate Texts in Mathematics*, Vol. 96, 2nd edn. Springer-Verlag, New York, 1990.
12. Tien-Yien Li, Finite approximation for the Frobenius-Perron operator. A solution to Ulam's conjecture. *Journal of Approximation Theory*, 1976, **17**, 177-186.
13. Schweitzer, P. J., Perturbation theory and finite Markov chains. *Journal of Applied Probability*, 1968, **5**(3), 401-404.
14. Haviv, M. and Van der Heyden, L., Perturbation bounds for the stationary probabilities of a finite Markov chain. *Advances in Applied Probability*, 1984, **16**, 804-818.
15. Seneta, E., Perturbation of the stationary distribution measured by ergodicity coefficients. *Advances in Applied Probability*, 1988, **20**, 228-230.
16. Kemeny, G. and Snell, J. L., *Finite Markov Chains.* Van Nostrand, New York, 1960.
17. Meyn, S. P. and Tweedie, R. L., *Markov Chains and Stochastic Stability.* Communications and Control Engineering. Springer-Verlag, London, 1993.
18. Meyer, C. D., The character of a finite Markov chain. *The IMA Volumes in Mathematics and its Applications* Vol. 48. Springer-Verlag, New York, 1993, pp. 47-58.
19. Seneta, E., Sensitivity of finite Markov chains under perturbation. *Statistics and Probability Letters*, 1993, **17**, 163-168.
20. Boyarsky, A. and Haddad, G., All invariant densities of piecewise linear Markov maps are piecewise constant. *Advances in Applied Mathematics*, 1981, **2**(3), 284-289.
21. Lasota, A. and Yorke, J., On the existence of invariant measures for piecewise monotonic transformations. *Transactions of the American Mathematical Society*, 1973, **186**, 481-488.
22. Keller, G., On the rate of convergence to equilibrium in one-dimensional systems. *Communications in Mathematical Physics*, 1984, **96**, 181-193.
23. Rychlik, M., Bounded variation and invariant measures. *Studia Mathematica*, 1983, **76**, 69-80.
24. Hofbauer, F. and Keller, G., Ergodic properties of invariant measures for piecewise monotonic transformations. *Mathematische Zeitschrift*, 1982, **180**, 119-140.
25. Keller, G., Stochastic stability in some chaotic dynamical systems. *Monatshefte für Mathematik*, 1982, **94**, 313-353.
26. Hunt, F. Y. and Miller, W. M., On the approximation of invariant measures. *Journal of Statistical Physics*, 1992, **66**(1/2), 535-548.