Theory Probab. Appl.
Vol. 47, No. 2

CONTRACTIVITY AND ERGODICITY OF THE RANDOM MAP $x \mapsto |x - \theta|$*

J. C. MATTINGLY†

Abstract. The long time behavior of the random map $x_n \mapsto x_{n+1} = |x_n - \theta_n|$ is studied under various assumptions on the distribution of the θ_n. One of the interesting features of this random dynamical system is that for a single fixed deterministic θ the map is not a contraction, while the composition is almost surely a contraction if θ is chosen randomly with only mild assumptions on the distribution of the θ's. The system is useful as an explicit model where more abstract ideas can be explored concretely. We explore various measures of convergence rates, hyperbolically from randomness, and the structure of the random attractor.

Key words. random dynamical systems, random attractors, random fix points, mixing, ergodicity

PII. S0040585X97979767

Consider the map $f_\theta\colon x \mapsto |x-\theta|$ as a map from $[0,1] \to [0,1]$ parameterized by $\theta \in [0,1]$. This paper explores the ergodic and contractive properties of $f_{\theta_n} \circ \cdots \circ f_{\theta_1}(x)$, where the θ_i's are independent, identically distributed random variables drawn according to common probability measure $\mathbf{P}$.

Let Ω denote the space of one-sided sequences $\{\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)\colon \theta_i \in [0,1]\}$. We endow Ω with the product measure generated by $\mathbf{P}$. We will also use $\mathbf{P}$ to denote this product measure. $\mathbf{E}$ will signify the expectation with respect to $\mathbf{P}$. If $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \dots)$ and $\tilde{\boldsymbol{\theta}} = (\theta_2, \theta_3, \dots)$, let σ denote the shift map on Ω defined by $\sigma\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$. σ^n will represent the n-fold composition of σ. Lastly, we define $f^n_{\boldsymbol{\theta}}(x) = f_{\theta_n} \circ \cdots \circ f_{\theta_1}(x)$.

We will explore the dynamics of $f^n_{\boldsymbol{\theta}}(x)$ by investigating the relative motion of two different initial conditions $X_0^{(1)}$ and $X_0^{(2)}$ subject to the same realization of randomness $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$. For the remainder of the paper, $X_0^{(i)}$ will denote some initial condition and we will set $X_n^{(i)} = f^n_{\boldsymbol{\theta}}(X_0^{(i)}) = f_{\sigma^{n-1}\boldsymbol{\theta}}(X_{n-1}^{(i)})$. We will neglect the superscript when considering a single trajectory.

At times we will also consider the dynamics obtained by starting with some initial condition at time $-n$ and evolving to time zero. To this end, we define Ω^* to be the space of two-sided infinite sequences $\{\boldsymbol{\theta} = (\dots, \theta_{-2}, \theta_{-1}, \theta_0, \theta_1, \theta_2, \dots)\colon \theta_i \in [0,1]\}$. The shift σ and $\mathbf{P}$ are extended to Ω^* in the natural way. We also define the backward iterates $f^{-n}_{\boldsymbol{\theta}}(x) = f_{\theta_0} \circ f_{\theta_{-1}} \circ \cdots \circ f_{\theta_{-n+1}}(x)$. As the randomness is composed in a different order, the dynamics given by $f^{-n}_{\boldsymbol{\theta}}(x)$ and $f^n_{\boldsymbol{\theta}}(x)$ are not equivalent a priori.

An interesting feature of this model is that even though $f_\theta \circ \cdots \circ f_\theta(x)$ is not contracting for a fixed θ, the composition $f_{\theta_n} \circ \cdots \circ f_{\theta_1}(x)$ with random θ_i's is contracting almost surely for a wide class of θ distributions. The random composition of otherwise noncontracting systems produces a contraction. Letac [10] seems to have been the first to conjecture that the backward iterates should contract for θ distributions with support which did not form a lattice. For an overview of these types of iterated systems see [6], [9]. After the completion of the investigation described in this paper the author became aware of a work [1] which addresses similar questions about the same systems. That analysis is organized slightly differently than this article but shares many aspects with it. In particular, the analysis

*Received by the editors November 22, 2001. This work was supported by NSF grant DMS-9971087.
http://www.siam.org/journals/tvp/47-2/97976.html

†Department of Mathematics, Stanford University, Stanford, CA (jonm@math.stanford.edu).

of (1) covers the case of when the law of θ is supported on only two points and makes more explicit how the mixing estimates depend on the arithmetic properties of the support of θ.

1. Ergodicity and main results. A probability measure μ is *invariant* (for the dynamics) if for any bounded measurable test function $\phi\colon [0,1] \to \mathbf{R}$,

$$\int \mathbf{E}\phi\big(f_{\boldsymbol\theta}(x)\big)\, d\mu(x) = \int \phi(x)\, d\mu(x).$$

LEMMA 1. *$f_{\boldsymbol\theta}$ possesses at least one invariant measure.*

Proof. This is a simple consequence of the compactness of $[0,1]$ and is obtained by the standard construction. One considers the empirical measures $\mu_N = N^{-1}\sum_{n=1}^{N} \delta_{X_n}$. Here δ_x is the delta measure concentrated on x. The collection $\{\mu_N\}$ is relatively compact and hence has limit points. Any limit point can be shown to be an invariant measure. See [9] for more details.

We now state the main result of this paper which will be proved in a number of steps in the subsequent sections.

THEOREM 2. *Consider the following two conditions*:

(1) *There exists a positive function $\Lambda\colon (0,1) \to (0,1)$ such that for any $\varepsilon > 0$*

$$\inf_{x\in[0,1-\varepsilon]} \mathbf{P}\big\{\theta \in [x, x+\varepsilon]\big\} \geqq \Lambda(\varepsilon) > 0;$$

(2) *there exist points α_1, α_2, and α_3 such that $\alpha_1 < \alpha_2 < \alpha_3$, α_1, and $\alpha_3 - \alpha_2$ are not rationally related, and $\mathbf{P}\{\theta \in [\alpha_i - \varepsilon, \alpha_i + \varepsilon]\} > 0$ for any $\varepsilon > 0$ and $i = 1, 2, 3$.*

If either condition above holds, then

(1) *the integrated random map $f_{\boldsymbol\theta}$ has a unique invariant measure*;

(2) *$f^n_{\boldsymbol\theta}$ and $f^{-n}_{\boldsymbol\theta}$ converge to a single-valued function. More exactly, for almost every $\boldsymbol\theta$*

$$\lim_{n\to\infty} \sup_{x,y\in[0,1]} |f^n_{\boldsymbol\theta}(x) - f^n_{\boldsymbol\theta}(y)| = 0 \quad \textit{and} \quad \lim_{n\to\infty} \sup_{x,y\in[0,1]} \left|f^{-n}_{\boldsymbol\theta}(x) - f^{-n}_{\boldsymbol\theta}(y)\right| = 0.$$

The statement of the above theorem is a bit redundant. The first condition of course implies the second. However, many common examples satisfy the first condition. Since the proof using it is much simpler, we separate the two conditions for expository reasons.

Although the above conditions are quite weak, there is numerical evidence that they are not optimal. It seems that if one has only α_1 and α_2 such that α_1, α_2, and 1 are not rationally related, then the system is ergodic. This is not proved here; however, the author believes that with a bit more care it could be proved using similar techniques.

2. A numerical interlude. Before embarking on the rigorous analysis, we share with the reader some numerical simulations the author found useful in developing intuition. We consider four different θ distributions: UNIFORM$[0,1]$, UNIFORM$\{e^{-1}, \frac{1}{2}, \pi/4\}$, UNIFORM$\{e^{-1}, \frac{3}{4}\}$, and UNIFORM$\{\pi/8, \pi/4\}$, where UNIFORM A denotes the uniform distribution on the set A. The first example has a positive probability of entering any set on the next iteration. The second and third examples do not, yet because the measure charges an incommensurate set of points one might hope that the system could still reach all configurations. Since all combinations of shifts in the last example live on a lattice, one does not expect it to be ergodic. The theorems here apply to the first two examples. Numerics suggest that with a little more care the same lines of reasoning could be applied to the third example. The paper [1] in fact shows the uniqueness of the invariant measure in this setting.

The plots on the left in Figure 1 give the position at $t = 0$ for 20 evenly spaced initial points starting at the given time in the past. All of the trajectories in a given plot use the same realization of noise. The plots on the right give the maximum separation at $t = 0$ for the collection of trajectories shown on the left.

What is interesting to note is that as we move the initial conditions farther and farther into the past in the first three cases, the value at $t = 0$ converges to a fixed value independent

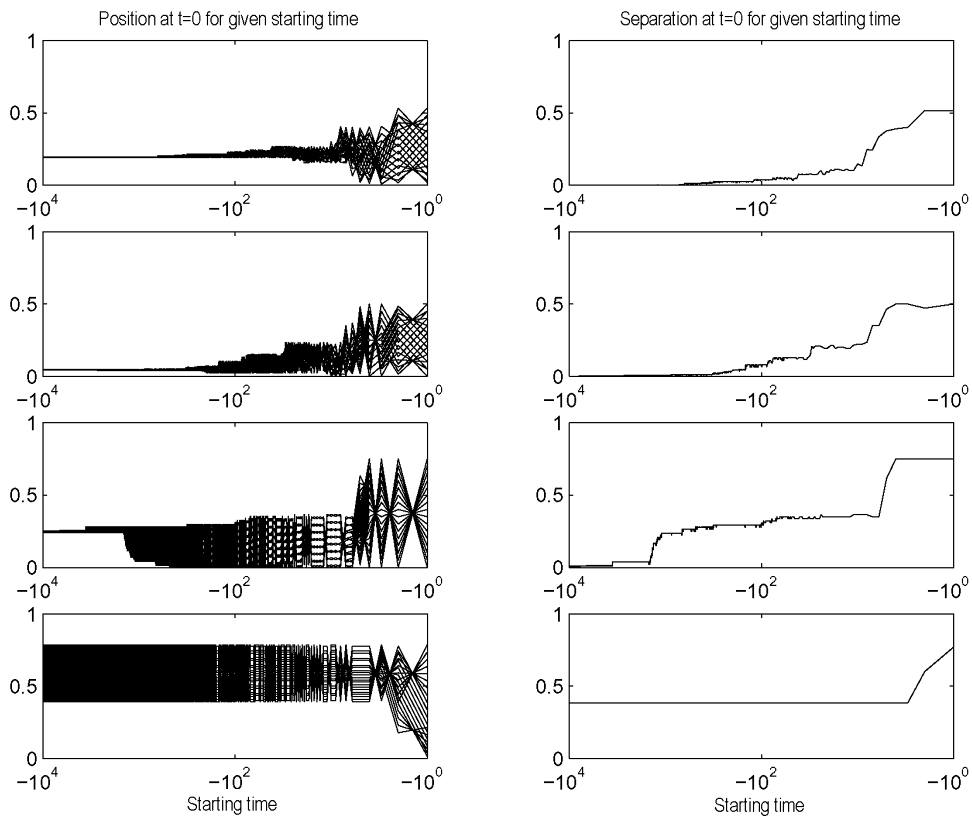

FIG. 1. *From top to bottom* θ *was distributed as* UNIFORM$[0,1]$, UNIFORM$\{e^{-1}, \frac{1}{2}, \pi/4\}$, UNIFORM$\{e^{-1}, \frac{3}{4}\}$, *and* UNIFORM$\{\pi/8, \pi/4\}$. *Notice that the bottom right plot is not on a log-log scale while the other plots on the left are.*

of the initial value. This complete loss of memory of the initial condition for almost every path leads to a strong form of ergodicity. The systems possess a random attractor consisting of a single point which attracts all initial data for a given realization of noise. This implies that the n-point motions are ergodic, not only the one-point motions. See [5], [3], and the references therein for more on random attractors.

The bottom pair of plots shows that the system is clearly not ergodic. Closer examination reveals that none of the trajectories have collapsed. Yet one sees that there is a subset of the domain into which all the trajectories are attracted. In this simple system this domain does not fluctuate as it does in more complicated systems.

We now turn to describing analytically the dynamics exhibited by the numerics 1.

3. Some soft theorems.

A few preliminary definitions are necessary. For any pair of initial conditions $X_0^{(1)}$ and $X_0^{(2)}$ define the sequence of stopping times

$$\tau_k\big(X_0^{(1)}, X_0^{(2)}\big) = \inf\left\{n\colon\ \frac{1}{2}\,|X_n^{(1)} - X_n^{(2)}| \leqq 2^{-k}\right\}.$$

For any collection of initial data $A \subset [0,1]$, define $\tau_k(A) = \sup\{\tau_k(x,y)\colon\ x, y \in A\}$. Lastly, we denote by $\mathrm{Lip}_1[0,1]$ the space of functions $\{f \in C([0,1], \mathbf{R})\colon\ |f(x) - f(y)| \leqq |x - y|\}$.

Assumption A.1. *For any k and any pair of initial conditions $X_0^{(1)}$ and $X_0^{(2)}$, $\tau_k(X_0^{(1)}, X_0^{(2)})$ is finite almost surely.*

As the following theorems show, this assumption is sufficient to guarantee the desired results with time running forward.

THEOREM 3. *Under Assumption* A.1,

$$\lim_{n\to\infty}\left|\mathbf{E}\phi(f^n_{\boldsymbol\theta}(X_0^{(1)}))-\mathbf{E}\phi(f^n_{\boldsymbol\theta}(X_0^{(2)}))\right|=0$$

for any $\phi\in \mathrm{Lip}_1[0,1]$.

This norm on measure obtained by the use of test functions from Lip_1 is usually called either the Wasserstein distance or the Kantorovich distance. If one uses test functions which are simply bounded and measurable one obtains the total variation distance on the space of measures. From this characterization one sees that convergence in the Wasserstein norm is weaker than convergence in total variation. Nonetheless the Wasserstein distance is a complete metric in this setting, hence guaranteeing the uniqueness of the limit points. See [7] for more details.

COROLLARY 1. *Under Assumption* A.1, *the iterated random map* $f_{\boldsymbol\theta}$ *has a unique invariant measure.*

Proof. Let μ and $\widetilde{\mu}$ be two invariant measures and let ϕ be an arbitrary bounded function in $\mathrm{Lip}_1[0,1]$. It is enough to show that $\int\phi(x)\,d\mu(x)=\int\phi(x)\,d\widetilde{\mu}(x)$ to conclude that $\mu=\widetilde{\mu}$ since bounded functions in Lip_1 characterize the probability measures on $[0,1]$. See [7].

Observe that

$$\begin{aligned}\left|\int\phi(x)\,d\mu(x)-\int\phi(\widetilde{x})\,d\widetilde{\mu}(\widetilde{x})\right|&=\left|\int\mathbf{E}\phi(f^n_{\boldsymbol\theta}(x))\,d\mu(x)-\int\mathbf{E}\phi(f^n_{\boldsymbol\theta}(\widetilde{x}))\,d\widetilde{\mu}(\widetilde{x})\right|\\&\leqq\int\int\mathbf{E}\left|\phi(f^n_{\boldsymbol\theta}(x))-\phi(f^n_{\boldsymbol\theta}(\widetilde{x}))\right|\,d\mu(x)\,d\widetilde{\mu}(\widetilde{x}).\end{aligned}$$

Since the integrand is bounded and positive, we conclude

$$\left|\int\phi(x)\,d\mu(x)-\int\phi(\widetilde{x})\,d\widetilde{\mu}(\widetilde{x})\right|\leqq\int\int\mathbf{E}\lim_{n\to\infty}\left|\phi\big(f^n_{\boldsymbol\theta}(x)\big)-\phi\big(f^n_{\boldsymbol\theta}(\widetilde{x})\big)\right|\,d\mu(x)\,d\widetilde{\mu}(\widetilde{x})\leqq 0.$$

Proof of Theorem 3. Let $\mathbf{1}_A(\boldsymbol\theta)$ denote the indicator function on the set A. Without loss of generality, we assume $\sup_x|\phi(x)|\leqq 1$. For any k,

$$\begin{aligned}\mathbf{E}\phi\big(f^n_{\boldsymbol\theta}(X_0^{(1)})\big)&=\mathbf{E}\phi\big(f^n_{\boldsymbol\theta}(X_0^{(1)})\big)\mathbf{1}_{\{\tau_k\leqq n\}}(\boldsymbol\theta)+\mathbf{E}\phi\big(f^n_{\boldsymbol\theta}(X_0^{(1)})\big)\mathbf{1}_{\{\tau_k>n\}}(\boldsymbol\theta)\\&=\mathbf{E}\phi\big(f^n_{\boldsymbol\theta}(X_0^{(1)})\big)\,\mathbf{1}_{\{\tau_k\leqq n\}}(\boldsymbol\theta)-\mathbf{E}\phi\big(f^n_{\boldsymbol\theta}(X_0^{(2)})\big)\,\mathbf{1}_{\{\tau_k\leqq n\}}(\boldsymbol\theta)\\&\quad+\mathbf{E}\phi\big(f^n_{\boldsymbol\theta}(X_0^{(2)})\big)\,\mathbf{1}_{\{\tau_k\leqq n\}}(\boldsymbol\theta)+\mathbf{E}\phi\big(f^n_{\boldsymbol\theta}(X_0^{(1)})\big)\,\mathbf{1}_{\{\tau_k>n\}}(\boldsymbol\theta)\\&\leqq\mathbf{E}\left|\phi\big(f^n_{\boldsymbol\theta}(X_0^{(1)})\big)-\phi\big(f^n_{\boldsymbol\theta}(X_0^{(2)})\big)\right|\,\mathbf{1}_{\{\tau_k\leqq n\}}(\boldsymbol\theta)+\mathbf{E}\phi\big(f^n_{\boldsymbol\theta}(X_0^{(2)})\big)+\mathbf{E}\mathbf{1}_{\{\tau_k>n\}}.\end{aligned}$$

Hence, by exchanging the roles of $X_0^{(1)}$ and $X_0^{(2)}$, by the definition of τ_k, and by the fact that ϕ is in Lip_1, we obtain

$$\left|\mathbf{E}\phi\big(f^n_{\boldsymbol\theta}(X_0^{(1)})\big)-\mathbf{E}\phi\big(f^n_{\boldsymbol\theta}(X_0^{(2)})\big)\right|\leqq 2^{-k}+\mathbf{P}\{\tau_k>n\}. \tag{1}$$

Since τ_k was assumed to be almost surely finite, $\lim_{n\to\infty}\mathbf{P}\{\tau_k>n\}=0$, implying that

$$\lim_{n\to\infty}\left|\mathbf{E}\phi\big(f^n_{\boldsymbol\theta}(X_0^{(1)})\big)-\mathbf{E}\phi\big(f^n_{\boldsymbol\theta}(X_0^{(2)})\big)\right|\leqq 2^{-k}.$$

As k was arbitrary, the proof is concluded.

We can in fact prove the following almost sure statement which highlights the contractive nature of the $f^n_{\boldsymbol\theta}$.

THEOREM 4. *If Assumption* A.1 *holds, then with probability one,*

$$\lim_{n\to\infty} \sup_{x,y\in[0,1]} \left| f_{\boldsymbol\theta}^n(x) - f_{\boldsymbol\theta}^n(y) \right| = 0.$$

Proof. Fix a k. Let $\{X_0^{(i)}\}_{i=1}^M$ be a partition of $[0,1]$ with $|X_0^{(i)} - X_0^{(i+1)}| \leqq 2^{-k}$. Since M is finite, $\tau_k(\{X_0^{(i)}\})$ is finite almost surely.

Because the graph of $f_{\boldsymbol\theta}^n$ is made up of segments with slopes of magnitude one, $|f_{\boldsymbol\theta}^n(z) - f_{\boldsymbol\theta}^n(y)| \leqq |z-y|$ for any $z, y \in [0,1]$ and any n. Fix z and y. Let $X_0^{(i)}$ and $X_0^{(j)}$ be the partition points closest to z and y, respectively. Notice that $|z - X_0^{(i)}|$ and $|y - X_0^{(j)}|$ are less than $2^{-(k+1)}$ by construction and thus, by the previous observation, $|f_{\boldsymbol\theta}^n(z) - f_{\boldsymbol\theta}^n(X_0^{(i)})|$ and $|f_{\boldsymbol\theta}^n(y) - f_{\boldsymbol\theta}^n(X_0^{(i)})|$ are less than $2^{-(k+1)}$. Hence, if $n > \tau_k(\{X_0^{(i)}\})$, then

$$\begin{aligned} \left| f_{\boldsymbol\theta}^n(z) - f_{\boldsymbol\theta}^n(y) \right| &\leqq \left| f_{\boldsymbol\theta}^n(z) - f_{\boldsymbol\theta}^n(X_0^{(i)}) \right| + \left| f_{\boldsymbol\theta}^n(X_0^{(i)}) - f_{\boldsymbol\theta}^n(X_0^{(j)}) \right| + \left| f_{\boldsymbol\theta}^n(X_0^{(j)}) - f_{\boldsymbol\theta}^n(y) \right| \\ &\leqq 2^{-(k+1)} + 2^{-(k+1)} + 2^{-(k+1)} = 3 \cdot 2^{-(k+1)}. \end{aligned}$$

Since x and y were arbitrary, we have $\sup_{z,y\in[0,1]} |f_{\boldsymbol\theta}^n(z) - f_{\boldsymbol\theta}^n(y)| \leqq 3 \cdot 2^{-(k+1)}$. Because $\tau_k(\{X_0^{(i)}\}) < \infty$ almost surely,

$$\lim_{n\to\infty} \sup_{z,y\in[0,1]} |f_{\boldsymbol\theta}^n(z) - f_{\boldsymbol\theta}^n(y)| \leqq 3 \cdot 2^{-(k+1)}$$

almost surely. As k was arbitrary, taking a countable sequence of bounds tending to zero proves the claim.

3.1. Limits from the distant past. We now consider limits obtained by starting with some initial condition at time $-n$ and evolving to time zero. We set $X_{-n,0}^{(i)} = f_{\boldsymbol\theta}^{-n}(X_0^{(i)})$. This is the value at time zero of dynamics starting from $X_0^{(i)}$ at time $-n$ with two-sided noise realization $\boldsymbol\theta \in \Omega^*$. The supposition is that as one starts further in the past, the trajectory settles down to a fixed value at time zero. In other words, we expect that the value at time zero has no memory of the initial condition used at $-\infty$. That is to say, it depends only on the realization of noise $\boldsymbol\theta$. To prove such a statement we use the following slightly stronger assumption.

Assumption A.2. *For any k and any pair of initial conditions $X_0^{(1)}$ and $X_0^{(2)}$, $\mathbf{E}[\tau_k(X_0^{(1)}, X_0^{(2)}]^p$ is finite for some $p > 1$.*

THEOREM 5. *Under Assumption* A.2, *there exists a random variable X^* measurable with respect to the randomness up to time zero such that*

$$\lim_{n\to\infty} \sup_{x\in[0,1]} \left| f_{\boldsymbol\theta}^{-n}(x) - X^*(\boldsymbol\theta) \right| = 0$$

almost surely. In addition, $X^*(\boldsymbol\theta)$ is stationary with respect to the shift σ (i.e., $X^*(\boldsymbol\theta)$ has the same distribution as $X^*(\sigma\boldsymbol\theta)$). Furthermore, $X^*(\boldsymbol\theta)$ is skew-invariant. That is to say, $f_{\boldsymbol\theta}(X^*(\boldsymbol\theta)) = X^*(\sigma\boldsymbol\theta)$ almost surely.

From Theorem 5 it follows that a unique invariant is the expected value of the measure $\delta_{X^*(\boldsymbol\theta)}(x) \times d\mathbf{P}(\boldsymbol\theta)$ which is invariant with respect to the skew-flow $(x, \boldsymbol\theta) \mapsto (f_{\boldsymbol\theta}(x), \sigma(\boldsymbol\theta))$. See [9] and [2] for more on this point of view.

Proof of Theorem 4. We proceed as in the proof of Theorem 4. First fix a k. Then choose a partition $\{X_0^{(i)}\}_{i=1}^M$ of $[0,1]$ with $|X_0^{(i)} - X_0^{(i+1)}| \leqq 2^{-k}$. Define $\tau_k^{(-n)}(\{X_0^{(i)}\})$ as before except start the initial data $\{X_0^{(i)}\}$ at time $-n$ and measure τ_k from that point in time. Clearly $\mathbf{E}(\tau_k^{(-n)}(\{X_0^{(i)}\}))^p$ is independent of n since the $\boldsymbol\theta$'s are stationary under the shift σ. By our assumption, we know it is finite for some $p > 1$.

Define the event $A_n = \{\boldsymbol\theta \in \Omega^* \colon \tau_k^{(-n)}(\{X_0^{(i)}\}) < n\}$. By Chebyshev's inequality, we have $\mathbf{P}(A_n) \leqq C(k) n^{-p}$. Since $p > 1$, the sum $\sum_n \mathbf{P}(A_n) < \infty$. Thus, the first Borel–Cantelli lemma (see [4]) and the argument contained in Theorem 4 about points not in the partition imply that there exists an $n_k^*(\boldsymbol{\theta})$ so that

$$n > n_k^*(\boldsymbol{\theta}) \Longrightarrow \sup_{x,y\in[0,1]} \left| f_{\boldsymbol{\theta}}^{-n}(x) - f_{\boldsymbol{\theta}}^{-n}(y) \right| \leqq 3 \cdot 2^{-(k+1)}.$$

Since k was arbitrary it is clear that in the limit the difference is zero. Furthermore, the limit $\lim_{n\to\infty} f_{\boldsymbol{\theta}}^{-n}(x)$ exists and is independent of x. Set $X^*(\boldsymbol{\theta}) = \lim f_{\boldsymbol{\theta}}^{-n}(1)$. Clearly, by construction, $X^*(\boldsymbol{\theta})$ is stationary, skew-invariant, and possesses the stated measurability.

4. The needed estimates. We have reduced the problem to finding conditions on the distributions of the θ_i's so that Assumptions A.1 and A.2 hold. We begin by making a number of observations about the dynamics of the two-point motion.

As before, we set $X_n^{(i)} = f_{\boldsymbol{\theta}}^n(X_0^{(i)}) = f_{\sigma^{n-1}\boldsymbol{\theta}}(X_{n-1}^{(i)})$ for initial conditions $X_0^{(1)}$ and $X_0^{(2)}$ and set noise realization $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$. Notice that if θ_n is not in between $X_{n-1}^{(1)}$ and $X_{n-1}^{(2)}$, then $|X_n^{(1)} - X_n^{(2)}| = |X_{n-1}^{(1)} - X_{n-1}^{(2)}|$. Otherwise the points move closer together. In particular, the distance between the two trajectories never increases.

To be more precise, let us define $z_n = \frac{1}{2}(X_n^{(1)} + X_n^{(2)})$ and $\rho_n = \frac{1}{2}|X_n^{(1)} - X_n^{(2)}|$. The (z_n, ρ_n) dynamics is equivalent to the $(X_n^{(1)}, X_n^{(2)})$ dynamics. The (z_n, ρ_n) evolves according to the following rules:

$$(z_{n+1}, \rho_{n+1}) = \begin{cases} (f_{\sigma^n\boldsymbol{\theta}}(z_n), \rho_n) & \text{if} \quad \theta_{n+1} \notin (z_n - \rho_n, z_n + \rho_n), \\ (\rho_n, f_{\sigma^n\boldsymbol{\theta}}(z_n)) & \text{if} \quad \theta_{n+1} \in (z_n - \rho_n, z_n + \rho_n). \end{cases}$$

There are a number of important features of this representation of the dynamics. First, if $\theta_{n+1} \notin (z_n - \rho_n, z_n + \rho_n)$, then $z_n \mapsto z_{n+1}$ moves as an ordinary point would under the dynamics. Second, if $\theta_{n+1} \in (z_n - \rho_n, z_n + \rho_n)$, then ρ_{n+1} is precisely the distance from θ_{n+1} to z_n. Hence, when θ_{n+1} comes close to z_n the points contract a lot.

4.1. Dense $\boldsymbol{\theta}$ distributions. We begin with a simple example which ensures Assumptions A.1 and A.2. This condition will be superseded by those in the next section, but it is easier to understand and contains most of the central ideas without many technical complications.

LEMMA 6. *If there exists a positive function* $\Lambda\colon (0,1) \to (0,1)$ *such that for any* $\varepsilon > 0$

$$\inf_{x\in[0,1-\varepsilon]} \mathbf{P}\left\{\theta \in [x, x+\varepsilon]\right\} \geqq \Lambda(\varepsilon) > 0,$$

then for any $k > 0$ *and initial* $X_0^{(1)}$ *and* $X_0^{(2)}$

$$\mathbf{P}\left\{\tau_k(X_0^{(1)}, X_0^{(2)}) > N\right\} \leqq \left[1 - \Lambda(2^{-k})\right]^N$$

which implies that Assumptions A.1 *and* A.2 *hold.*

Observe that it is sufficient that the measure $\mathbf{P}$ have a positive continuous density with respect to Lebesgue measure to satisfy the assumption of Lemma 6.

Proof of Lemma 6. If at any time $\theta_{n+1} \in [z_n - 2^{-k-1}, z_n + 2^{-k-1}]$, then $\rho_{n+1} \leqq 2^{-k-1}$ which implies that $|X_{n+1}^{(1)} - X_{n+1}^{(2)}| \leqq 2^{-k}$ as desired. By assumption, the probability that $\theta_{n+1} \notin [z_n - 2^{-k-1}, z_n + 2^{-k-1}]$ is less than $1 - \Lambda(2^{-k})$. Since all the θ_i's are independent, we have the quoted estimate.

4.2. More minimal conditions on the distribution of $\boldsymbol{\theta}$. The assumption made in the previous section is quite reasonable. It is satisfied by many natural examples such as the uniform distribution. Nonetheless, it is natural to ask: What are the minimal conditions to ensure Assumptions A.1 and A.2? This section takes steps to answer that question.

LEMMA 7. *Assume that there exist points* α_1, α_2, *and* α_3 *such that* $\alpha_1 < \alpha_2 < \alpha_3$, α_1, *and* $\alpha_3 - \alpha_2$ *are not rationally related, and* $\mathbf{P}\{\theta \in [\alpha_i - \varepsilon, \alpha_i + \varepsilon]\} > 0$ *for any* $\varepsilon > 0$ *and*

$i = 1, 2, 3$. *Then Assumptions* A.1 *and* A.2 *hold. In fact, for any fixed k there exist an $r \in (0,1)$ and positive C so that*

$$\mathbf{P}\Big\{\tau_k(X_0^{(1)}, X_0^{(2)}) \geqq N\Big\} \leqq Cr^N.$$

Before we prove Lemma 3 we first prove a simpler version which will be used to prove Lemma 7.

LEMMA 8. *Assume that there exist points α_1, α_2, and α_3 such that $\alpha_1 < \alpha_2 < \alpha_3$, α_1, and $\alpha_3 - \alpha_2$ are not rationally related, and $\mathbf{P}\{\theta = \alpha_i\} > 0$ for $i = 1, 2, 3$. Then Assumptions* A.1 *and* A.2 *hold. In fact, for any fixed k there exist an $r \in (0,1)$ and positive C so that*

$$\mathbf{P}\Big\{\tau_k(X_0^{(1)}, X_0^{(2)}) \geqq N\Big\} \leqq Cr^N.$$

We will prove this in a number of steps. For $z_0 \in [0,1]$, $\varepsilon_1 > \varepsilon_2 > 0$, and integer N, we introduce the events

$$B(z_0, \varepsilon_1, \varepsilon_2, N) = \Big\{\boldsymbol{\theta} \in \Omega\colon\ \exists\, k,\ k \leqq N,\ \text{such that } \theta_k \in [z_{k-1} - \varepsilon_2, z_{k-1} + \varepsilon_2] \text{ and } \theta_j \notin [z_{j-1} - \varepsilon_1, z_{j-1} + \varepsilon_1] \text{ for } 0 \leqq j < k\Big\},$$

where $z_n = f_{\boldsymbol{\theta}}^n(z_0)$. The usefulness of these events stems from two facts. Consider the setting where $\varepsilon_1 = \frac{1}{2}|X_0^{(1)} - X_0^{(2)}|$. First, since $\theta_j \notin [z_{j-1} - \varepsilon_1,\, z_{j-1} + \varepsilon_1]$ for $0 \leqq j < k$, we know that $z_j = \frac{1}{2}(X_j^{(1)} + X_j^{(2)})$ for $j < k$. Second, as $\theta_k \in [z_{k-1} - \varepsilon_2, z_{k-1} + \varepsilon_2]$, we know that $\rho_k = \frac{1}{2}|X_j^{(1)} - X_j^{(2)}| \leqq \varepsilon_2$.

LEMMA 9. *Fix any $\alpha_1 < \alpha_2 < \alpha_3$ as in Lemma* 8. *Then there is an ε_0 such that if $0 < \varepsilon_2 < \varepsilon_1 \leqq \varepsilon_0$, then for a $z_0 \in [0,1]$ there exist an N and a $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$ with the following properties*:

(i) $\theta_i \in \{\alpha_1, \alpha_2, \alpha_3\}$ *for all* $i = 1, \dots, N$;

(ii) $\theta_N = \alpha_1$ *and* $|f_{\boldsymbol{\theta}}^{N-1}(z_0) - \alpha_1| \leqq \varepsilon_2$;

(iii) $|f_{\boldsymbol{\theta}}^{k-1}(z_0) - \theta_k| > \varepsilon_1$ *for* $k = 1, \dots, N-2$.

This implies that in the setting of Lemma 8, $\mathbf{P}\{B(z_0, \varepsilon_1, \varepsilon_2, N')\} > 0$ *for any* $N' \geqq N$.

Proof. First, notice that if $x \leqq \alpha_2$, then $f_{\alpha_3} \circ f_{\alpha_2}(x) = x + \alpha_3 - \alpha_2$. Setting $T(x) = x + \alpha_3 - \alpha_2$, observe that because α_1 is incommensurate with $\alpha_3 - \alpha_2$, the orbit of $T \mod \alpha_1$ is dense in the interval $[0, \alpha_1]$. Hence, there exists some M so that $T^M(z_0) \mod \alpha_1$ is within ε_2 of α_1. Fix this M. We now simply need to translate this into a sequence of α_i's which satisfy the other constraints of the lemma. We do this by judiciously inserting applications of f_{α_1} to keep the trajectory below α_2, where the action of $f_{\alpha_3} \circ f_{\alpha_2}(x)$ equals that of T.

We set $\varepsilon_0 = (\alpha_2 - \alpha_1)/2$ and assume $\varepsilon_2 < \varepsilon_1 \leqq \varepsilon_0$. We use the following algorithm to build $\boldsymbol{\theta}$.

- INITIALIZE: set $m = 0$ and $N = 0$.
- WHILE $(m < M)$ DO:
  - If $z_N \in (\alpha_1 + \varepsilon, 1]$, then set $\theta_{N+1} = \alpha_1$, $z_{N+1} = f_{\alpha_1}(z_N)$, and $N = N + 1$.
  - If $z_N \in [0, \alpha_1 + \varepsilon]$, then set $\theta_{N+1} = \alpha_2$, $\theta_{N+2} = \alpha_3$, $z_{N+1} = f_{\alpha_2}(z_N)$, $z_{N+2} = f_{\alpha_3}(z_{N+1})$, $M = M + 1$, $N = N + 2$.
- Set $\theta_{N+1} = \alpha_1$, $z_{N+1} = f_{\alpha_1}(z_N)$, and $N = N + 1$.
- OUTPUT: The $(\theta_1, \theta_2, \dots, \theta_N)$ constructed has the desired properties. N gives the length of the needed trajectory.

Notice that by construction N is finite. Since in the setting of Lemma 8 the probability of θ equaling any given α_i is positive, the chance of taking the constructed path is strictly positive. This concludes the proof.

We now extend Lemma 9 to be uniform over initial conditions.

LEMMA 10. *Fix any $\alpha_1 < \alpha_2 < \alpha_3$ as in Lemma 8. There exists an ε_0 such that for any $0 < \varepsilon_2 < \varepsilon_1 < \varepsilon_0$ there are an N and a $\gamma > 0$ such that*

$$\inf_{z\in[0,1]} \mathbf{P}\Big\{B(z,\varepsilon_1,\varepsilon_2,N)\Big\} \geqq \gamma > 0.$$

Proof. Set $\eta = \frac{1}{3}\min(\varepsilon_2, \varepsilon_0 - \varepsilon_1)$, $\varepsilon_1' = \varepsilon_1 + \eta$, and $\varepsilon_2' = \varepsilon_2 - \eta$. Now fix some z_0 and fix the N constructed in Lemma 9 using ε_1' and ε_2'. By the construction of the trajectory in Lemma 8 and the choices of ε_1' and ε_2', the event $B(z_0,\varepsilon_1',\varepsilon_2',N)$ implies the event $B(z_0',\varepsilon_1,\varepsilon_2,N)$ for any $z_0' \in [0,1]$ with $|z_0' - z_0| \leqq \eta$. Hence, $\mathbf{P}\{B(z_0',\varepsilon_1,\varepsilon_2,N)\} \geqq \mathbf{P}\{B(z_0,\varepsilon_1',\varepsilon_2',N)\} > 0$ for any such z_0'.

Now consider a partition $\{z_0^{(i)}\}_{i=1}^M$ of $[0,1]$ with maximum spacing η. From Lemma 9 we know there exist $N^{(i)}$ and positive $\gamma^{(i)}$ such that if $N' \geqq N^{(i)}$, then $\mathbf{P}\{B(z_0^{(i)},\varepsilon_1',\varepsilon_2',N')\} \geqq \gamma^{(i)} > 0$. Now set $N = \max N^{(i)}$ and $\gamma = \min \gamma^{(i)}$. Since there are only finitely many points $y_0^{(i)}$, γ is strictly positive and N is finite. Because for any $x \in [0,1]$, $|x - y_0^{(i)}| \leqq \eta$ for some i the observations in the preceding paragraph imply that $\inf_{x\in[0,1]} \mathbf{P}\{B(x,\varepsilon_1,\varepsilon_2,N)\} \geqq \gamma > 0$.

Proof of Lemma 8. It is enough to prove the claim when $|X_0^{(1)} - X_0^{(2)}| < \varepsilon_0$, where ε_0 is given by Lemma 10. When the separation is larger, we can use a finite sequence of intermediary points to link the two and then require that all these points come sufficiently close to one another to ensure that the original trajectories of interest are within 2^{-k}.

We set $\varepsilon_1 = |X_0^{(1)} - X_0^{(2)}|$, $\varepsilon_2 = 2^{-k}$, and $z_0 = \frac{1}{2}(X_0^{(1)} + X_0^{(2)})$ and find the N and γ guaranteed by Lemma 10. The event $B(z_0,\varepsilon_1,\varepsilon_2,N)$ implies that $|X_N^{(1)} - X_N^{(2)}| \leqq 2^{-k}$. Furthermore, we know that $\mathbf{P}\{B(z_{jN},\varepsilon_1,\varepsilon_2,N)\} \geqq \gamma > 0$ is independent of j because the estimate is uniform over the initial point. Hence, we have the estimate $\mathbf{P}\{\tau_k(X_0^{(1)},X_0^{(2)}) > jN\} \leqq (1-\gamma)^j$ which implies the claim.

Proof of Lemma 7. This lemma will be proved by the same means as Lemma 8. The only difference is that we must build our desired trajectory in Lemma 9 out of a sequence of theta's from small neighborhoods of α_1, α_2, and α_3. We start by finding a sequence of alpha's that would work corresponding to a slightly larger ε_1 and a slightly smaller ε_2. Knowing the length of this sequence we can choose small enough neighborhoods about the α_i to ensure that if we replace all the α_i's with any element of the neighborhoods about them, the sequence will still produce a trajectory with the desired properties relative to the original ε_1 and ε_2. In short, we use the previous results coupled with the continuity of $f_{\boldsymbol{\theta}}^N(x)$ with respect to $\boldsymbol{\theta}$.

5. Modes and rates of convergence. We have shown that under fairly mild conditions, $x \mapsto |x-\theta|$ has a unique invariant measure which attracts in the Wasserstein distance any initial distribution. We also showed that the system possesses a trivial random attractor. In fact, we showed that if the system starts at $-\infty$, it converges to a single value at time zero which depends only on the realization of the noise. In section 3, it was shown that contractivity of the map implies the uniqueness of the invariant measure. However, it is clear that it is not necessary. In fact the rate of the convergence to the invariant measure induced by the dynamics and the rate of almost sure contraction can be quite different.

To explore these issues we consider the simplest of cases, namely when θ is uniformly distributed on $[0,1]$. It is straightforward to verify that $2(1-x)\,dx$ is a unique invariant measure for this system. We begin by examining the two-point contraction rate.

In the notation of the previous sections, the only way that $|X_n^{(1)} - X_n^{(2)}| > \varepsilon$ (assuming $|X_0^{(1)} - X_0^{(2)}| > \varepsilon$) is if none of the θ_i's, $i = 1,\dots,n$, fall in the ε-neighborhood of $z_i = \frac{1}{2}(X_i^{(1)} + X_i^{(2)})$. The chance of this event is independent of the position of z_i; in fact,

$$\mathbf{P}\left\{|X_n^{(1)} - X_n^{(2)}| > \varepsilon\right\} = (1-\varepsilon)^n. \tag{2}$$

Hence if we want the probability for separation to be small (say, $e^{-\gamma}$ for some $\gamma > 0$), then we need to take $\gamma/|\log(1-\varepsilon)|$ steps. When ε is small this means that n is essentially γ/ε.

This fact implies a bound on the convergence to the invariant measure in the Wasserstein norm. The estimate in (1) implies that the Wasserstein distance after $n = \log(1/\varepsilon)/\varepsilon$ steps is at most 2ε. Hence we see that, except for a logarithmic correction, this bound suggests that the distance decays like $1/n$. Even in this explicit case this estimate need not give the sharp rate of convergence of the empirical distribution to the stationary distribution. Numerically it is true that the distance between two points subject to the same realization of noise does decay like $1/n$. This is seen in (2). However, if one is interested only in the convergence of the induced measure, one is not constrained to use that same realization.

To see the difference, we will estimate the convergence rate to the invariant distribution in total variation norm directly without looking at the contractive properties of each realization. If $P^n(x, \cdot)$ is the distribution induced on $[0, 1]$ by moving n steps in the Markov chain starting from x, then direct calculation gives that

$$|P(x,\cdot) - P(y,\cdot)|_{TV} = \tfrac{1}{2}\,|y-x| - \tfrac{1}{4}\,|y-x|^2 \geqq \tfrac{1}{2}. \tag{3}$$

Thus the chain satisfies the classical Döblin condition which produces the estimate $|P(x,\cdot) - P(y,\cdot)|_{TV} \leqq 2^{-n}$. Since the total variation norm dominates the Wasserstein norm, we see that the distribution of the random variable approaches its equilibrium distribution exponentially quickly.

Which of these it the "right" answer? It of course depends on the question. If the map f models some process which transports mass around the interval, then the rate at which the many-point motion contracts to a single point is quite relevant. If f models the trajectory of a single system, the fact that its distribution becomes randomized exponentially quickly might be of greater interest.

6. Conclusions.

This paper illustrates one mechanism which produces ergodicity; namely, almost sure contraction. This behavior is far from universal, but when it exists, it yields a lot of information about the process. It is interesting to note that the random iterated map is a contraction while $f_\theta \circ f_\theta \circ \cdots \circ f_\theta$ is not a contraction for fixed constant θ.

The use of three points in the second condition of Theorem 2 is likely unnecessary. With more care similar arguments should give the same results with only two points, under appropriate conditions. Numerical experiments support this opinion.

In general, convergence rates can also be obtained in this framework. One uses estimates on the moments of the stopping times τ_k. Some uniformity of the estimates in k is needed. This is straightforward under the first condition of Theorem 2. However, it requires more work in the other setting and the author has not attempted this.

In section 5, we showed that considering the contraction rate does not always give the correct rate of convergence to the stationary distribution. However, it does have its advantages. It is particularly useful when the natural topology of the invariant measure is difficult to predict. This is not the case when the distribution of θ is absolutely continuous with respect to Lebesgue measure. The topology is less clear when the distribution of θ is atomic; however, it is not unsurmountable. There are settings where the difficulties are much more challenging. When the phase space is infinite dimensional, as in stochastic PDEs or particle systems, there are many nonequivalent topologies. In these cases, the approach detailed in this paper has succeeded when others have failed. (See [11], [8] for examples.)

Acknowledgments. The author thanks Persi Diaconis for asking the question which led to this paper, for then listening to the answers, and generally providing engaging discussions. The author also thanks Arron Abrams, whose request for the unpublished notes which preceded this paper pushed the author to finally finish this document.

REFERENCES

[1] A. Abrams, H. Landau, Z. Landau, J. Pommersheim, and E. Zaslov, *An iterated random function with Lipschitz number one*, Theory Probab. Appl., 47 (2002), pp. 190–201.

[2] P. H. Baxendale, *Statistical equilibrium and two-point motion for a stochastic flow of diffeomorphisms*, Progr. Probab., 19 (1991), pp. 189–218.

[3] P. H. Baxendale, *Stability and equilibrium properties of stochastic flows of diffeomorphisms*, Progr. Probab., 27 (1992), pp. 3–35.

[4] P. Billingsley, *Probability and Measure*, Wiley, Chichester, 1995.

[5] H. Crauel and F. Flandoli, *Attractors for random dynamical systems*, Probab. Theory Related Fields, 100 (1994), pp. 365–393.

[6] P. Diaconis and D. Freedman, *Iterated random functions*, SIAM Rev., 41 (1999), pp. 45–76.

[7] R. M. Dudley, *Probabilities and Metrics. Convergence of Laws on Metric Spaces, with a View to Statistical Testing*, Lecture Notes Ser. 45, Matematisk Institut, Aarhus Universitet, Aarhus, 1976.

[8] W. E, K. Khanin, A. Mazel, and Ya. Sinai, *Invariant measures for Burgers equation with stochastic forcing*, Ann. of Math. (2), 151 (2000), pp. 877–960.

[9] Yu. Kifer, *Ergodic Theory of Random Transformations*, Birkhäuser, Boston, 1986.

[10] G. Letac, *A contraction principle for certain Markov chains and its applications*, Contemp. Math., 50 (1986), pp. 263–273.

[11] J. C. Mattingly, *Ergodicity of 2D Navier–Stokes equations with random forcing and large viscosity*, Comm. Math. Phys., 206 (1999), pp. 273–288.