Journal of Theoretical Probability, Vol. 1, No. 3, 1988

Random Walks Generated by Affine Mappings

Marc A. Berger[1,2] **and H. Mete Soner**[2]

Received June 19, 1987

This article is concerned with Markov chains on $\mathbb{R}^m$ constructed by randomly choosing an affine map at each stage, and then making the transition from the current point to its image under this map. The distribution of the random affine map can depend on the current point (i.e., state of the chain). Sufficient conditions are given under which this chain is ergodic.

KEY WORDS: Ergodic; Markov chain; random affine map.

1. INTRODUCTION

Let G be a set of affine maps from $\mathbb{R}^m$ into itself. To each $x \in \mathbb{R}^m$ associate a probability measure μ_x on G. Construct a Markov chain $\{X_n\}$ on $\mathbb{R}^m$ according to the recursion $X_{n+1} = g_{n+1} X_n$, $n \geqslant 0$, where $g_{n+1} \in G$ is distributed according to μ_{X_n}, independent of $X_0, \ldots, X_{n-1}$. This chain has transition probabilities $\mathbb{P}(x, A) = \mu_x(gx \in A)$. The larger chain $\{(g_n, X_n)\}$ on $G \times \mathbb{R}^m$ is referred to as the *embedded* chain.

The article in Ref. 2 was concerned with properties of its invariant distribution, when the chain $\{X_n\}$ is ergodic. Here conditions involving logarithmic rather than pth moments are given, under which the chain is ergodic. More significant in contrast to Ref. 2 is the introduction here of Khas'minskii's construction of an invariant measure.[5,6] This construction is based on the sampled chain $\{X_{\tau^k}\}$ at its return times τ^k to some compact subset of $\mathbb{R}^m$. As long as there is average contractivity under the probabilities μ_x these return times have finite mean, and the invariant measure can be normalized to an invariant probability.

A discussion of the random matrix setting, where G is a set of invertible $m \times m$ matrices, appears in Furstenberg and Kifer.[4] More generally if

[1] Faculty of Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel.
[2] Department of Mathematics, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.

0894-9840/88/0700-0239$06.00/0

860/1/3-1

the affine maps here in G have a common invariant compact set, then Furstenberg and Kifer's setting of a random walk on a compact metric space applies. For the scalar case $m = 1$ this article is a continuation along the lines of Brandt[3] and Vervaat.[7] These authors consider the stochastic difference equation

$$X_{n+1} = a_{n+1} X_n + b_{n+1}, \qquad n \geqslant 0$$

In Ref. 7 the sequence of pairs $\{(a_n, b_n)\}$ is taken to be i.i.d., and in Ref. 3 it is taken to be stationary ergodic. Brandt and Vervaat give conditions for the existence and uniqueness of stationary solutions $\{X_n\}$. Some of the present work is ased on Barnsley *et al.*,[1] who consider the case where G is a finite set of maps. In this connection the first author would like to acknowledge the helpful conversation and correspondence he had with J. Elton.

Notation. The standard notation $\mathbb{P}_x$, $\mathbb{E}_x$, Var_x is used to indicate probability and statistics with respect to the probability measure μ_x. The one-step transition operator T defined on $C_b(\mathbb{R}^m)$ is given by $Tf(x) = \mathbb{E}_x f(X_1)$. For $g \in G$ write $g\colon x \mapsto ax + b$, letting a generically denote the linear part and b the shift part of the map. In this way μ_x also induces measures on the sets of $m \times m$ matrices and m vectors.

2. ERGODICITY

The assumptions of the probabilities μ_x necessary for the arguments below as follows:

(A1) Strong continuity of measures. For any $h \in C_b(G)$ and $\in \mathbb{R}^m$

$$|\mathbb{E}_x h - \mathbb{E}_y h| \leqslant \varphi(|x - y|) \, \|h\|$$

where $\varphi\colon \mathbb{R}_+ — \mathbb{R}_+$ is a continous function and

$$\lim_{t \downarrow 0} \varphi(t) \log^2 \frac{1}{t} = 0$$

(A2) Uniform Average Contractivity. There exist $0 < \varepsilon_0 < 1$, 0 such that

$$\sup_{x, |y|} \mathbb{E}_x[\log |ay| \,\big|\, |ay| \geqslant \varepsilon_0] \leqslant -\alpha$$

where, as above, a denotes a linear part of g.

(A3) Uniform positivity of contraction probability. For every $\varepsilon > 0$ there exist $0 < s < 1$ such that

$$\inf_{|x-x'| \geqslant \varepsilon} |gx - g'x'| \leqslant s|x-x'|) := \beta > 0$$

where g and g' are independently distributed according to μ_x and $\mu_{x'}$, respectively.

(A4) Integrability. There exist $0 < \delta \leqslant 1$, $K > 0$ such that

$$\sup_{x, |y|} \mathbb{E}_x(\log^+ |ay|)^{2+\delta} \leqslant K$$

$$\sup_{x} \mathbb{E}_x(\log^+ |b|)^{1+\delta} \leqslant K$$

where, as above, b denotes the shift part of g.

Theorem. Under assumptions (A1)–(A4) $\{X_n\}$ is ergodic, and has a unique stationary asymptotic distribution.

The proof of this result relies on the following lemmas.

Lemma 1. Under assumptions (A2), (A4) for any $r > \max(\varepsilon_0, e^{-\alpha})$ there exists $R > 1$ such that

$$\sup_x \mathbb{E}_x \log^+(|ay| + |b|) \leqslant \log(r|y|) \tag{2.1}$$

whenever $|y| \geqslant R$.

Proof. Define $\tilde{a}$: $\mathbb{R}^m \to \mathbb{R}^m$ by

$$\tilde{a}(y) := \begin{cases} ay & \text{if} \quad |ay| \geqslant \varepsilon_0 |y| \\ \varepsilon_0 y & \text{if} \quad |ay| < \varepsilon_0 |y| \end{cases} \tag{2.2}$$

Observe that although it is not necessarily a linear map, $\tilde{a}$ is homogeneous of degree one. Moreover, for any $y \in \mathbb{R}^m$

$$|\tilde{a}(y)| \geqslant \max(|ay|, \varepsilon_0 |y|) \tag{2.3}$$

An analogue of assumption (A2) holds for $\tilde{a}$; namely,

$$\sup_{x, |y|=1} \mathbb{E}_x \log |\tilde{a}(y)| \leqslant -\min\left(\alpha, \log \frac{1}{\varepsilon_0}\right) := -\tilde{\alpha} < 0 \tag{2.4}$$

Since $r > \max(\varepsilon_0, e^{-\alpha})$ is follows that $\log r > -\tilde{\alpha}$. Choose $\varepsilon_1 > 0$ and $R > 1$ in succession so that

$$\log(1+\varepsilon_1) = \frac{\log r + \tilde{\alpha}}{2} \tag{2.5}$$

$$\log(\varepsilon_0 \varepsilon_1 R) \geqslant \max\left\{\log\left(\frac{\varepsilon_1}{1+\varepsilon_1}\right), K^{1/(1+\delta)}\left[2\,\frac{\log(1+1/\varepsilon_1) + K^{1/(1+\delta)})}{\log r + \tilde{\alpha}}\right]^{1/\delta}\right\} \tag{2.6}$$

Let $x, y \in \mathbb{R}^m$ with $|y| \geqslant R$. Then

$$\begin{aligned}
&\int_{|\tilde{a}(y)| \leqslant (1/\varepsilon_1)|b|} \log^+(|\tilde{a}(y)| + |b|)\, \mu_x(dg) \\
&\quad \leqslant \int_{|\tilde{a}(y)| \leqslant (1/\varepsilon_1)|b|} \log^+\left[\left(1 + \frac{1}{\varepsilon_1}\right)|b|\right] \mu_x(dg) \\
&\quad \leqslant \mathbb{E}_x^{1/(1+\delta)}\left[\log\left(1+\frac{1}{\varepsilon_1}\right) + \log^+ |b|\right]^{1+\delta} \mathbb{P}_x^{\delta/(1+\delta)}(|b| \geqslant \varepsilon_0 \varepsilon_1 R) \\
&\quad \leqslant \left[\log\left(1+\frac{1}{\varepsilon_1}\right) + K^{1/(1+\delta)}\right] \frac{K^{\delta/(1+\delta)}}{\log^\delta(\varepsilon_0 \varepsilon_1 R)} \leqslant \frac{\log r + \tilde{\alpha}}{2}
\end{aligned}$$

The second inequality above follows from Hölder's inequality and (2.3). The third inequality follows from the triangle and Chebyshev inequalities; and finally the fourth inequality uses (2.6).

One can also estimate the other integral

$$\begin{aligned}
&\int_{|\tilde{a}(y)| > (1/\varepsilon_1)|b|} \log^+[|\tilde{a}(y)| + |b|]\, \mu_x(dg) \\
&\quad \leqslant \int_{|\tilde{a}(y)| > (1/\varepsilon_1)|b|} \log^+[(1+\varepsilon_1)|\tilde{a}(y)|]\, \mu_x(dg) \\
&\quad \leqslant \int_{(1+\varepsilon_1)|\tilde{a}(y)| > 1} \log[(1+\varepsilon_1)|\tilde{a}(y)|]\, \mu_x(dg) = \mathbb{E}_x \log[(1+\varepsilon_1)\,|\tilde{a}(y)|] \\
&\quad \leqslant -\tilde{\alpha} + \log|y| + \log(1+\varepsilon_1) = \log|y| + \frac{\log r - \tilde{\alpha}}{2}
\end{aligned}$$

The first equality above follows from (2.3) and (2.6). The third inequality follows from (2.4) since $\tilde{a}$ is homogeneous of degree one, and the second equality follows from (2.5). Combining these last two estimates and using (2.3) now leads to (2.1). □

Remark. The proof of Lemma 1 only uses the second part of assumption (A4), involving b. The first part of this assumption (alone), involving a, is used in the proof of the next lemma.

Lemma 2. Under assumptions (A1), (A2), (A4) for any $f \in C_0(\mathbb{R}^m)$, $\{T^n f\colon n \geqslant 0\}$ is an equicontinuous sequence of functions.

Proof. Let $\tilde{a}$ be defined again as in (2.2) above. Along with (2.4) an analogue of assumption (A4) holds for $\tilde{a}$ as well, namely,

$$\sup_{x,\, |y|=1} \mathbb{E}_x \left| \log |\tilde{a}(y)| \right|^{2+\delta} \leqslant K + 2 \log^{2+\delta} \frac{1}{\varepsilon_0} := \tilde{K} \tag{2.7}$$

Define

$$\psi(t) := \frac{1}{\log(1/t)} \wedge 1, \qquad t > 0$$

On account of assumption (A1)

$$\lim_{t \downarrow 0} \frac{\varphi(t)}{\psi(t) - \psi(e^{-\tilde{\alpha}} t)} = 0$$

where $\tilde{\alpha} = \min(\alpha, \log(1/\varepsilon_0))$ was defined above in (2.4). Thus for

$$C := (8 + 2^{2+\delta})\, \tilde{K} \tag{2.8}$$

it also holds that

$$\lim_{t \downarrow 0} \frac{\varphi(t)}{\psi(t) - \psi(e^{-\tilde{\alpha}} t(- C\psi^{2+\delta}(t)} = 0$$

since this extra term is of lower order $o(1/\log^2 t)$. Furthermore for $t > 0$ sufficiently small

$$\psi(t) > \psi(e^{-\tilde{\alpha}} t) + C\psi^{2+\delta}(t)$$

Thus it follows from the continuity of φ that for $0 < t_0 < e^{-2}$ sufficiently small

$$\varphi(t) \leqslant \psi(t) - \psi(e^{-\tilde{\alpha}} t) - C\psi^{2+\delta}(t), \qquad 0 < t \leqslant t_0 \tag{2.9}$$

Observe that $\psi(t) \geqslant t$, $0 < t < e^{-1}$. Suppose that $f \in C_0(\mathbb{R}^m)$ is in fact Lipschitz continuous. Then by scaling f if necessary it may be assumed that $\|f\| \leqslant \psi(t_0)/2$, and that $|f(x)-f(y)| \leqslant \psi(|x-y|)$ for all $x, y \in \mathbb{R}^m$. From this it follows by induction (as demonstrated below) that

$$|T^k f(x) - T^k f(y)| \leqslant \psi(|x-y|) \tag{2.10}$$

for all $k \geqslant 0$ and all $x, y \in \mathbb{R}^m$. To argue this suppose that (2.10) holds for $k = n-1$ and all $x, y \in \mathbb{R}^m$. Then by assumption (A1) using the fact that $\|T^n f\| \leqslant \psi(t_0)/2 < 1$; for $x, y \in \mathbb{R}^m$

$$\begin{aligned} |T^n f(x) - T^n f(y)| &\leqslant \mathbb{E}_x \, ||T^{n-1} f(gx) - T^{n-1} f(gy)| \\ &\quad + |\mathbb{E}_x T^{n-1} f(gy) - \mathbb{E}_y T^{n-1} f(gy)| \\ &\leqslant \mathbb{E}_x \psi(|a(x-y)|) + \varphi(|x-y|) \end{aligned} \tag{2.11}$$

Fix $z \in \mathbb{R}^m$, $0 < |z| \leqslant t_0$. Since $|z| \leqslant e^{-2}$ it follows that

$$\tilde{\psi}(r) := \psi(e^r) = -1/r, \qquad -\infty < < r \leqslant \tfrac{1}{2} \log |z|$$

Then one checks that

$$\tilde{\psi}(r) - \frac{8}{|\log^3 |z||} r^2$$

is concave, $-\infty < r \leqslant \frac{1}{2} \log |z|$. This is used now to estimate $\int_{|\tilde{a}(z)| \leqslant |z|^{1/2}} \psi(|\tilde{a}(z)|) \, \mu_x(dg)$. First introduce the notation

$$S := \{|\tilde{a}(z)| \leqslant |z|^{1/2}\}, \qquad Y := \log |\tilde{a}(z)|$$

Then by Jensen's inequality

$$\begin{aligned} \int_{|\tilde{a}(z)| \leqslant |z|^{1/2}} \psi(|\tilde{a}(z)|) \, \mu_x(dg) &= \mathbb{P}_x(S) \, \mathbb{E}_x[\psi(|\tilde{a}(z)|) \,|\, S] \\ &= \mathbb{P}_x(S) \, \mathbb{E}_x[\tilde{\psi}(Y) \,|\, S] \\ &\leqslant \mathbb{P}_x(S) \left\{ \tilde{\psi}(\mathbb{E}_x[Y \,|\, S]) + \frac{8}{|\log^2 |z||} \mathrm{Var}_x[Y \,|\, S] \right\} \end{aligned} \tag{2.12}$$

Since $\tilde{a}$ is homogeneous of degree one, the term $\mathrm{Var}_x[Y \,|\, S] = \mathrm{Var}_x[\log |\tilde{a}(z)| \,|\, S]$ depends only on the *direction* of z, but not on its magnetitude. Thus if $\hat{z} := z/|z|$, then

$$\mathrm{Var}_x[Y \,|\, S] = \mathrm{Var}_x[\log |\tilde{a}(\hat{z})| \,|\, S] \leqslant \frac{\mathbb{E}_x \log^2 |\tilde{a}(\hat{z})|}{\mathbb{P}_x(S)} \leqslant \frac{\tilde{K}}{\mathbb{P}_x(S)} \tag{2.13}$$

using (2.7) in this last step. To estimate the term $\mathbb{E}_x[Y \mid S]$ observe that by (2.4)

$$\begin{aligned}\int_S Y\mu_x(dg) &= \log|z|\ \mathbb{P}_x(S) + \int_S \log|\tilde{a}(\hat{z})|\ \mu_x(dg)\\ &\leqslant -\tilde{\alpha} + \log|z|\ \mathbb{P}_x(S) - \underbrace{\int_{|\tilde{a}(\hat{z})| > |z|^{-1/2}} \log|\tilde{a}(\hat{z})|\ \mu_x(dg)}_{\text{positive}}\\ &\leqslant -\tilde{\alpha} + \log|z|\ \mathbb{P}_x(S)\end{aligned}$$

Thus

$$\mathbb{E}_x[Y \mid S] \leqslant -\frac{\tilde{\alpha}}{\mathbb{P}_x(S)} + \log|z| \leqslant -\tilde{\alpha} + \log|z| \tag{2.14}$$

Plugging (2.13) and (2.14) into (2.12), using the fact that $\tilde{\psi}$ is increasing, leads to

$$\int_{|\tilde{a}(z)| \leqslant |z|^{1/2}} \psi(|\tilde{a}(z)|)\ \mu_x(dg) \leqslant \psi(e^{-\tilde{\alpha}}|z|) + 8\tilde{K}\psi^3(|z|) \tag{2.15}$$

The estimate of $\int_{|\tilde{a}(z)| > |z|^{1/2}} \psi(|\tilde{a}(z)|)\ \mu_x(dg)$ is straightforward, by use of Chebyshev's inequality:

$$\begin{aligned}\int_{|\tilde{a}(z)| > |z|^{1/2}} \psi(|\tilde{a}(z)|)\ \mu_x(dg) &\leqslant \mathbb{P}_x(|\tilde{a}(\hat{z})| > |z|^{-1/2})\\ &\leqslant \frac{2^{2+\delta}\mathbb{E}_x|\log|\tilde{a}(\hat{z})||^{2+\delta}}{|\log|z||^{2+\delta}} \leqslant 2^{2+\delta}\tilde{K}\psi^{2+\delta}(|z|)\end{aligned} \tag{2.16}$$

using the fact that $\|\psi\| = 1$ in the first step and (2.7) in the last step. Combining the two estimates (2.15) and (2.16), and using $\psi^3(|z|) \leqslant \psi^{2+\delta}(|z|)$ and (2.3) along with the fact that ψ is increasing, leads to the upper bound

$$\mathbb{E}_x\psi(|az|) \leqslant \psi(e^{-\tilde{\alpha}}|z|) + C\psi^{2+\delta}(|z|)$$

for $0 < |z| \leqslant t_0$, where C was defined above in (2.8). Thus, going back to (2.11) now, with $z = x - y$, we have

$$|T^n f(x) - T^n f(y)| \leqslant \varphi(|z|) + \psi(e^{-\tilde{\alpha}}|z|) + C\psi^{2+\delta}(|z|) \leqslant \psi(|z|)$$

for $0 < |z| \leqslant t_0$, using (2.9) in this last step.

This establishes (2.10) for $k=n$ in the case $|x-y| \leqslant t_0$. On the other hand if $|x-y| \geqslant t_0$ then (2.10) is obvious, since $\|T^k f\| \leqslant \psi(t_0)/2$ and ψ is increasing. Therefore this lemma is proved for the case where f is Lipschitz continuous. Since Lipschitz continuous functions are dense in $C_0(\mathbb{R}^m)$, it follows that $\{T^n f\colon n \geqslant 0\}$ is equicontinuous for any fixed $f \in C_0(\mathbb{R}^m)$. □

Lemma 3. Let $\{Y_n : n \geqslant 0\}$ be a real-valued non-negative stochastic process adapted relative to the filtration of σ-algebras $\{\mathscr{F}_n : n \geqslant 0\}$. Fix $\varepsilon > 0$ and define $\tau := \inf\{n \geqslant 0\colon Y_n \leqslant \varepsilon\}$. Assume the following:

(B1) There exists $0 < s < 1$, $\beta > 0$ such that

$$\mathbb{P}(Y_n \leqslant s Y_{n-1} \mid \mathscr{F}_{n-1}) \geqslant \beta I_{\{Y_{n-1} > \varepsilon\}}, \qquad \forall n \geqslant 1$$

(B2)

$$\lim_{M \uparrow \infty} \mathbb{P}(Y_n \leqslant M \text{ i.o.}) = 1$$

Then $\tau < \infty$ a.s.

Proof. Given any $\delta > 0$ it follows from (B2) that one can choose M so that $\mathbb{P}(Y_n \leqslant M \text{ i.o.}) \geqslant 1 - \delta$. For this choice then choose N such that $s^N M \leqslant \varepsilon$. Define stopping times $\gamma_0 := 0$,

$$\gamma_k := \inf\{n \geqslant \gamma_{k-1} + N\colon Y_n \leqslant M\}, \qquad k \geqslant 1$$

and set $A = \{\gamma_k = \infty \text{ for some } k\}$. On account of the way M was chosen, $\mathbb{P}(A) \leqslant \delta$.

Next set $B = \{\tau = \infty\} \backslash A$. If $\mathbb{P}(B) = 0$ then the desired conclusion of this lemma follows, since then $\mathbb{P}(\tau = \infty) \leqslant \mathbb{P}(A) \leqslant \delta$, and δ was arbitrary. So the remainder of this proof is devoted to establishing that $\mathbb{P}(B) = 0$.

Denote $\gamma_{k,L} := \gamma_k \wedge L$ and set

$$B_{k,L} := \{\gamma_{k,L} < L, \tau > \gamma_{k-1,L} + N\}, \qquad k \geqslant 1$$

Observe that by assumption (B1)

$$\mathbb{P}(\tau \leqslant \gamma_{k,L} + N \mid \mathscr{F}_{\gamma_{k,L}}) \geqslant \beta^N I_{\{\gamma_{k,L} < L, \tau > \gamma_{k,L}\}} \tag{2.17}$$

This is intuitively clear since $s^N M < \varepsilon$, and $Y_{\gamma_{k,L}} \leqslant M$ whenever $\gamma_{k,L} < L$; but to give a formal prof one needs to introduce some notation. Denote

$$C_i := \{Y_{\gamma_{k,L}+i} \leqslant s Y_{\gamma_{k,L}+i-1}\}, \qquad D_i := \{\tau > \gamma_{k,L} + i\}$$

$$E_i := \{Y_{\gamma_{k,L}+i} > \varepsilon\}, \qquad F := \{\gamma_{k,L} < L\}, \qquad \mathscr{G}_i := \mathscr{F}_{\gamma_{k,L}+i}$$

Assumption (B(1) implies that

$$\mathbb{P}(C_i \mid \mathscr{G}_{i-1}) \geqslant \beta I_{E_{i-1}} \tag{2.18}$$

and since $s^N M < \varepsilon$

$$\bigcap_{i=1}^{N} C_i \cap F \subset \tilde{D}_N \tag{2.19}$$

(the tilde denotes set complement). Thus by (2.19)

$$\mathbb{P}(\tilde{D}_N \mid \mathscr{G}_0) \geqslant \mathbb{P}(\tilde{D}_{N-1} \mid \mathscr{G}_0) + \mathbb{P}\left(\bigcap_{i=1}^{N} C_i \cap D_{N-1} \cap F \mid \mathscr{G}_0\right)$$

Furthermore using (2.18)

$$\begin{aligned}\mathbb{P}\left(\bigcap_{i=1}^{N} C_i \cap D_{N-1} \cap F \mid \mathscr{G}_0\right) &= \mathbb{E}[\mathbb{P}(C_N \mid \mathscr{G}_{N-1}) I_{\bigcap_{i=1}^{N-1} C_i \cap D_{N-1} \cap F} \mid \mathscr{G}_0] \\ &\geqslant \beta \mathbb{P}\left(\bigcap_{i=1}^{N-1} C_i \cap D_{N-1} \cap F \mid \mathscr{G}_0\right)\end{aligned}$$

This can be continued now as follows:

$$\begin{aligned}\mathbb{P}(\tilde{D}_N \mid \mathscr{G}_0) &\geqslant \mathbb{P}(\tilde{D}_{N-1} \mid \mathscr{G}_0) + \beta \mathbb{P}\left(\bigcap_{i=1}^{N-1} C_i \cap D_{N-1} \cap F \mid \mathscr{G}_0\right) \\ &\geqslant \mathbb{P}(\tilde{D}_{N-1} \mid \mathscr{G}_0) + \beta\left[\mathbb{P}\left(\bigcap_{i=1}^{N-1} C_i \cap D_{N-2} \cap F \mid \mathscr{G}_0\right) - \mathbb{P}(D_{N-2} \backslash D_{N-1} \mid \mathscr{G}_0)\right] \\ &\geqslant \mathbb{P}(\tilde{D}_{N-1} \mid \mathscr{G}_0) - \beta \mathbb{P}(D_{N-2} \backslash D_{N-1} \mid \mathscr{G}_0) + \beta^2 \mathbb{P}\left(\bigcap_{i=1}^{N-2} C_i \cap D_{N-2} \cap F \mid \mathscr{G}_0\right) \\ &\geqslant \cdots \\ &\geqslant \mathbb{P}(\tilde{D}_{N-1} \mid \mathscr{G}_0) - \beta \mathbb{P}(D_{N-2} \backslash D_{N-1} \mid \mathscr{G}_0) - \beta^2 \mathbb{P}(D_{N-3} \backslash D_{N-2} \mid \mathscr{G}_0) - \cdots \\ &\quad \beta^{N-1} \mathbb{P}(D_0 \backslash D_1 \mid \mathscr{G}_0) + \beta^N \mathbb{P}(D_0 \cap F \mid \mathscr{G}_0)\end{aligned}$$

Since $\beta \leqslant 1$

$$\begin{aligned}&\beta \mathbb{P}(D_{N-2} \backslash D_{N-1} ' \mathscr{G}_0) + \beta^2 \mathbb{P}(D_{N-3} \backslash D_{N-2} \mid \mathscr{G}_0) \\ &\quad + \cdots + \beta^{N-1} \mathbb{P}(D_0 \backslash D_1 \mid \mathscr{G}_0) \leqslant \mathbb{P}(\tilde{D}_{N-1} \mid \mathscr{G}_0)\end{aligned}$$

and so (2.17) follows.

On account of (2.17)

$$\begin{aligned}\mathbb{P}(B_{k,L(}\leqslant \mathbb{P}(\gamma_{k-1,L}<L,\tau>\gamma_{k-1,L}+N)\\ &= \mathbb{E}[\mathbb{P}(\tau>\gamma_{k-1,L}+N\mid \mathscr{F}_{\gamma_{k-1,L}})\, I_{\{\gamma_{k-1,L}<L\}}]\\ &\leqslant (1-\beta^N)\,\mathbb{P}(\tau>\gamma_{k-1,L},\gamma_{k-1,L}<L)\leqslant (1-\beta^N)\,\mathbb{P}(B_{k-1,L})\end{aligned}$$

One can let $L\uparrow\infty$ here and obtain

$$\mathbb{P}(B_k)\leqslant (1-\beta^N)\,\mathbb{P}(B_{k-1})$$

where

$$B_k := B_{k,\infty} = \{\gamma_k<\infty,\ \tau<\gamma_{k-1}+N\}$$

Thus

$$\mathbb{P}(B)=\mathbb{P}\left(\bigcap_{k=1}^{\infty} B_k\right)=\lim_{k\to\infty}\mathbb{P}(B_k)=0 \qquad \square$$

Let $\{X_n\}$ be the Markov chain described above in Section 1. Fix $R>1$ and define a sequence of stopping times $\{\tau^k: k\geqslant 0\}$ as follows:

$$\tau^0 := 0, \qquad r^{k+1} := \inf\{n\geqslant \tau^k+1: |X_n|\leqslant R\}, \qquad k\geqslant 0$$

Lemma 4. Under assumptions (A2), (A4), one can choose $R>1$ so that

$$\mathbb{E}_x\tau^1\leqslant D(1+\log^+|x|), \qquad \forall x\in\mathbb{R}^m \tag{2.20}$$

for some $D>0$.

Proof. Choose $e^{-\tilde{\alpha}}<r<1$, where $\tilde{\alpha}$ was defined above in (2.4), and pick $R>1$ so that)2.1) holds whenever $|y|\geqslant R$. Lemma 1 assures that this can be done. Let $\mathscr{F}_n$ denote the σ algebra generated by X_0 and g_i, $1\leqslant i\leqslant n$. The event $\{\tau^1\geqslant n+1\}$ belongs to $\mathscr{F}_n$. Thus by the Markov property and (2.1) we have

$$\begin{aligned}\int_{\tau^1\geqslant n+1}\log^+|X_{n+1}|\,\mu_x(dg) &= \int_{\tau^1\geqslant n+1}\mathbb{E}_x[\log^+|X_{n+1}|\,\big|\,\mathscr{F}_n]\,\mu_x(dg)\\ &= \int_{\tau^1\geqslant n+1}\mathbb{E}_{X_n}\log^+|aX_n+b|\,\mu_x(dg)\leqslant \int_{\tau^1\geqslant n+1}\log|rX_n|\,\mu_x(dg)\\ &\leqslant \log r\,\mathbb{P}_x(\tau^1\geqslant n+1)+\int_{\tau^1\geqslant n}\log^+|X_n|\,\mu_x(dg)\end{aligned}$$

The next-to-last inequality here makes use of the fact that $\{\tau^1 \geqslant n+1\}$ is contained in the event $\{|X_n| > R\}$, so that (2.1) applies. By iterating the above inequality on n one arrices at

$$\sum_{k=2}^{\infty} \mathbb{P}_x(\tau^1 \geqslant k) \leqslant \frac{\mathbb{E}_x \log^+ |X_1|}{\log(1/r)}$$

To estimate $\mathbb{E}_x \log^+ |X_1|$ observe that

$$\begin{aligned}\mathbb{E}_x \log^+ |X_1| &\leqslant \mathbb{E}_x \log(|ax| + |b| + 1) \leqslant \mathbb{E}_x \log(|ax| + 1) + \mathbb{E}_x \log(|b| + 1) \\ &\leqslant 2 \log 2 + \int_{|ax| \geqslant 1} \log(2|ax|)\, \mu_x(dg) + \int_{|b| \geqslant 1} \log(2|b|)\, \mu_x(dg) \\ &\leqslant 4 \log 2 + 2K + \log^+ |x|\end{aligned}$$

using assumption (A4) in this last step. Since $\mathbb{E}_x \tau^1 = \sum_{k=1}^{\infty} \mathbb{P}_x(\tau^1 \geqslant k)$ this establishes (2.20) for the choice

$$D = 1 + \frac{4 \log 2 + 2K}{\log(1/r)} \qquad \square$$

Now cnsider the sampled Markov chain $\{X_{\tau^k} : k \geqslant 0\}$. Since its state space $B_R = \{|x| \leqslant R\}$ is compact it must have at least one stationary measure, say $\bar{\nu} \in \mathcal{M}(B_R)$. Then

$$\int_{B_R} f(x)\, \bar{\nu}(dx) = \int_{B_R} \mathbb{E}_x f(X_{\tau^1})\, \bar{\nu}(dx), \qquad \forall f \in C(B_R) \tag{2.21}$$

For S a Borel subset of $\mathbb{R}^m$ let $N_*(S)$ be the number of times $X_k \in S$, $0 \leqslant k \leqslant \tau^1 - 1$. Define a Borel measure ν_* on $\mathbb{R}^m$ by

$$\nu_*(S) = \mathbb{E}_{\bar{\nu}} N_*(S)$$

Observe that for any $f \in C_b(\mathbb{R}^m)$ we have

$$\int_{\mathbb{R}^m} f(x)\, \nu_*(dx) = \int_{B_R} \mathbb{E}_x \left[\sum_{k=0}^{\tau^1 - 1} f(X_k) \right] \bar{\nu}(dx) \tag{2.22}$$

Lemma 5 (Khas'minskii). ν_* is a stationary measure for the chain $\{X_n\}$.

Proof. Let $f \in C_b(\mathbb{R}^m)$. Then by (2.22) and the Markov property

$$\int_{\mathbb{R}^m} \mathbb{E}_x f(X_1)\, \nu_*(dx) = \int_{B_R} \mathbb{E}_x \left[\sum_{k=0}^{\tau^1 - 1} \mathbb{E}_{X_k} f(X_1) \right] \bar{\nu}(dx)$$

$$= \int_{B_R} \mathbb{E}_x \left[\sum_{k=1}^{\tau^1} f(X_k) \right] \bar{\nu}(dx) = \int_{B_R} \mathbb{E}_x \left[\sum_{k=0}^{\tau^1 - 1} f(X_k) \right] \bar{\nu}(dx)$$

$$+ \int_{B_R} \mathbb{E}_x f(X_{\tau^1})\, \bar{\nu}(dx) - \int_{B_R} f(x)\, \bar{\nu}(dx) = \int_{\mathbb{R}^m} f(x)\, \nu_*(dx)$$

using (2.21) and (2.22) in this very last step. □

If R is chosen as in Lemma 4 then

$$\nu_*(\mathbb{R}^m) = \mathbb{E}_{\bar{\nu}} \tau^1 \leqslant D(1 + \log R) < \infty$$

since $\bar{\nu}$ is supported on B_R. Thus ν_* can be normalized to a stationary probability

$$\nu = \frac{\nu_*}{\mathbb{E}_{\bar{\nu}} \tau^1} \tag{2.23}$$

Proof of the Theorem. It suffices to show that for any $f \in C_0(\mathbb{R}^m)$, $x \in \mathbb{R}^m$

$$\lim_{n \to \infty} \mathbb{E}_x f(X_n) = \lim_{n \to \infty} T^n f(x) = \int_{\mathbb{R}^m} f(x)\, \nu(dx) \tag{2.24}$$

where ν is the stationary probability constructed above in (2.23). Fix $x \in \mathbb{R}^m$. Let $\{X_n\}$ be the chain with initial distribution ν, and let $\{X'_n\}$ be another chain independent of $\{X_n\}$ with $X'_0 = x$. Then for the embedded chains $\{(g_n, X_n)\}$ and $\{(g'_n, X'{}_n)\}$, g_n and g'_n are independent. Set $Y_n := |X_n - X'_n|$.

Let $\varepsilon > 0$ be arbitrary. It is first shown that assumptions (B1) and (B2) of Lemma 3 hold for $\{Y_n\}$ and the filtration of σ algebras $\{\mathscr{F}_n\}$, where $\mathscr{F}_n = \sigma(X_0, g_i, g'_i\colon\ 1 \leqslant i \leqslant n)$. On account of the Markov property and assumption (A3) there exists $0 < s < 1$ such that

$$\mathbb{P}(Y_n \leqslant s Y_{n-1} \,|\, \mathscr{F}_{n-1})$$
$$= \mathbb{P}(|g_n X_{n-1} - g'_n X'_{n-1}| \leqslant s |X_{n-1} - X'_{n-1}| \,\big|\, X_{n-1}, X'_{n-1}) \geqslant \beta I_{\{Y_{n-1} > \varepsilon\}}$$

This establishes (B1).

On account of Lemma 5 the process $\{X_n\}$ is stationary. Let $\{\tau^k : k \geqslant 0\}$ be the return times of $\{X'_n\}$ to the ball B_R, as in the discussion above. Since τ_k is independent of $\{X_n\}$

$$\mathbb{P}(|X_{\tau^k}| > M - R) = \nu(\mathbb{R}^m \backslash B_{M-R})$$

for any $M > R$, $k \geqslant 1$. Thus for $M > R$, since $|X'_{\tau^k}| \leqslant R$, $\forall k \geqslant 1$, it follows that

$$\begin{aligned}\mathbb{P}(|Y_n| \leqslant M \text{ i.o.}) &\geqslant \mathbb{P}(|X_{\tau^k}| \leqslant M - R \text{ i.o.})\\ &\geqslant 1 - \liminf_{k \to \infty} \mathbb{P}(|X_{\tau^k}| > M - R)\\ &= 1 - \nu(\mathbb{R}^m \backslash B_{M-R}) = \nu(B_{M-R})\end{aligned}$$

This establishes (B2). Thus the conclusion of Lemma 3 holds.

Fix $f \in C_0(\mathbb{R}^m)$ and let $\varepsilon' > 0$ be arbitrary. According to Lemma 2 $\{T^n f: n \geqslant 0\}$ is equicontinuous; so choose $\varepsilon > 0$ such that $|T^n f(x) - T^n f(y)| \leqslant \varepsilon'$, $\forall n \geqslant 0$, whenever $|x - y| \leqslant \varepsilon$. Corresponding to this choice of ε set $\tau := \inf\{n \geqslant 0\colon Y_n \leqslant \varepsilon\}$. Then by the Markov property

$$\begin{aligned}T^n f(x) - \int_{\mathbb{R}^m} f(x)\, \nu(dx) &= \mathbb{E}_x[f(X'_n) - f(X_n)]\\ &= \sum_{m=0}^{n} \int_{\tau = m} [T^{n-m} f(X'_m) - T^{n-m} f(X_m)] + \int_{\tau > n} [f(X'_n) - f(X_n)]\end{aligned}$$

Thus

$$\left| T^n f(x) - \int_{\mathbb{R}^m} f(x)\, \nu(dx) \right| \leqslant \varepsilon' + 2\, \|f\|\, \mathbb{P}(\tau > n)$$

Since $\tau < \infty$ a.s. by Lemma 3 and $\varepsilon' > 0$ was arbitrary, this leads to (2.24) and completes the proof. □

3. DISCUSSION OF ASSUMPTIONS (A1)–(A4)

Assumption (A2) clearly follows from the following stronger hypothesis:

(A2)' There exist $0 < \varepsilon_0 < 1$, $\alpha > 0$ such that

$$\sup_{x, |y| = 1} \int_{|ay| \geqslant \varepsilon_0} \log |ay|\, \mu_x(dg) \leqslant -\alpha$$

It is shown now that assumptions (A2)' and (A4) follow from the following stronger hypotheses:

(A2)'' There exists $\alpha > 0$ such that

$$\sup_{x, |y| = 1} \mathbb{E}_x \log |ay| \leqslant -\alpha$$

(A4)′ There exist $0<\delta\leqslant 1$, $K>0$ such that

$$\sup_{x,\,|y|=1} \mathbb{E}_x\,\big|\log|ay|\,\big|^{2+\delta}\leqslant K$$

$$\sup_x \mathbb{E}_x(\log^+|b|)^{1+\delta}\leqslant K$$

Clearly (A4) follows from (A4)′. Furthermore if (A4)′ holds then (A2)′ and (A2)″ are equivalent. Indeed for any $0<\varepsilon<1$ and $y\in\mathbb{R}^m$ with $|y|=1$

$$\begin{aligned}\mathbb{E}_x\log|ay| &\leqslant \int_{|ay|\geqslant\varepsilon}\log|ay|\;\mu_x(dg)=\mathbb{E}_x\log|ay|-\int_{|ay|<\varepsilon}\log|ay|\;\mu_x(dg)\\ &\leqslant -\alpha+\mathbb{E}_x^{1/(1+\delta)}\big|\log|ay|\,\big|^{1+\delta}\cdot\mathbb{P}_x^{1/(1+\delta)}(|ay|<\varepsilon)\\ &\leqslant -\alpha+K^{1/(1+\delta)}\left[\frac{K}{\log^{1+\delta}(1/\varepsilon)}\right]^{\delta/(1+\delta)}=-\alpha+\frac{K}{\log^{\delta}(1/\varepsilon)}\end{aligned}$$

using Hölder's and Chebyshev's inequalities and assumption (A4)′ in the last two inequalities. If $\varepsilon_0=\varepsilon$ is chosen small enough then this last term will be negative.

Assumption (A3) follows from the following stronger hypothesis:

(A3)′ There exists $0<s<1$ such that for any $\varepsilon>0$

$$\inf_{x,x'}\mathbb{P}(|gx-g'x'|\leqslant s|x-x'|+\varepsilon)>0$$

where g and g' are independently distributed according to μ_x and μ'_x, respectively.

Finally the example presented below shows that the strong continuity in assumption (A1) cannot be relaxed to weak continuity, if the chain $\{X_n\}$ is to be ergodic.

Example. Take $m=1$ and let μ_x be the atomic measure with

$$\alpha=1/2 \qquad \text{with prob } 1$$

$$b=\begin{cases}[-\tfrac{1}{2}(x\wedge 1)+(2x\wedge 1)]\vee 0 & \text{with prob } 3/4\\ 0 & \text{with prob } 1/4\end{cases}$$

Then under μ_x, for $0\leqslant x\leqslant 1$,

$$gx=\begin{cases}2x\wedge 1 & \text{with prob } 3/4\\ \tfrac{1}{2}x & \text{with prob } 1/4\end{cases}$$

Clearly these measures satisfy (A2)″, (A3)′, and (A4)′. For (A3)′ observe that

$$\mathbb{P}(|gx - g'x'| \leqslant \tfrac{1}{2}|x - x'|) \geqslant \tfrac{1}{16}$$

They are also weakly continuous, but there exist at least two stationary probabilities here:

$$\nu_1 = \delta_0$$

$$\nu_2 = \tfrac{2}{3} \sum_{k=0}^{\infty} 3^{-k} \delta_{2^{-k}}$$

It is clear that the deficiency is that b in this example contains part of the linear term that should be in a; and had this linear term been in a, the contractive assumption (A2) would have then failed to hold. One way to overcome this is to write the random map g as

$$g(\xi)\colon x \mapsto a(x, \xi)x + b(x, \xi)$$

where ξ is a random variable on some probability space taking values in some measurable space. Then it could be assumed that $a(x, \xi)$ and $b(x, \xi)$ are of order zero in x.

In general one can consider Markov chains on $\mathbb{R}^m$ which evolve according to

$$X_{n+1} = g(X_n, \xi_{n+1})$$

where $\{\xi_i\}$ is an i.i.d. sequence of random variables ξ_i: $(\Omega, \mathscr{F}, \mathbb{P}) \to (\Omega', \mathscr{F}')$, and $g\colon \mathbb{R}^m \times \Omega' \to \mathbb{R}^m$ is joint measurable. The setup above corresponds to the special case

$$g(x, \xi) = a(x, \xi)x + b(x, \xi)$$

where $a(x, \xi)$ is an $m \times m$ matrix and the following hold:

(C1) For any $h \in C_b(G)$, G being the space of affine maps $x \mapsto ax + b$, and for any $x, y \in \mathbb{R}^m$

$$|\mathbb{E}[h(a(x, \xi), b(x, \xi)) - h(a(y, \xi), b(y, \xi))]| \leqslant \varphi(|x - y|)\,\|h\|$$

where $\varphi\colon \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous function and

$$\lim_{t \downarrow 0} \varphi(t) \log^2 \frac{1}{t} = 0$$

(C2) There exist $0 < \varepsilon_0 < 1$, $\alpha > 0$ such that

$$\sup_{x, |y| = 1} \mathbb{E}[\log |a(x, \xi)\, y| \,\big|\, |a(x, \xi)\, y| \geqslant \varepsilon_0] \leqslant -\alpha$$

(C3) For any $\varepsilon > 0$ there exists $0 < s < 1$ such that

$$\inf_{|x - x'| \geqslant \varepsilon} \mathbb{P}(|g(x, \xi_1) - g(x', \xi_2)| \leqslant s\, |x - x'|) := \beta > 0$$

(C4) There exist $0 < \delta \leqslant 1$, $K > 0$ such that

$$\sup_{x, |y| = 1} \mathbb{E}[\log^+ |a(x, \xi)\, y|]^{2 + \delta} \leqslant K$$

$$\sup_x \mathbb{E}[\log^+ |b(x, \xi)|]^{1 + \delta} \leqslant K$$

ACKNOWLEDGMENT

Research was suported by AFOSR grant No. AFOSR-87-0137.

REFERENCES

1. Barnsley, M. F., Demko, S. G., Elton, J. H., and Geronimo, J. S. (1988). Invariant measures for Markov process arising from iterated function systems with place-dependent probabilities, Georgia Tech. preprint.
2. Berger, M. A., and Amit, Y. (1988). Products of random affine maps (to appear).
3. Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Prob.* **18**, 211–220.
4. Furstenberg, H., and Kifer, Y. (1983). Random matrix products and measures on projective space. *Israel J. Math.* **46**, 12–32.
5. Khas'minskii, R. Z. (1960). Ergodic properties of recurrent diffusion processes and stabilization of the solution to the Cauchy problem for parabolic equations. *Theor. Prob. Appl.* **5**, 179–196.
6. Khas'minskii, R. Z. (1980). *Stochastic Stability of Differential Equations*, Sijthoff and Noordhoff, Rockville, Maryland (translated from Russian).
7. Vervaat, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. Appl. Prob.* **11**, 750–783.