JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **138**, 232–248 (1989)

Stochastic Perturbation of Dynamical Systems: The Weak Convergence of Measures

ANDRZEJ LASOTA

Institute of Mathematics, University of Marii Curie-Sklodowskiej, 20-031 Lublin, Poland

AND

MICHAEL C. MACKEY

Department of Physiology, McGill University, Montreal, Quebec, Canada H3G 1Y6

Submitted by C. Foias

Received April 13, 1987

In this paper we consider the asymptotic behaviour of randomly perturbed discrete dynamical systems. We treat this problem by examining the evolution of the corresponding sequences of distributions. We show that an average contractive property is sufficient to ensure the weak convergence of the sequence of distributions to a unique stationary measure.

1. INTRODUCTION

The study of stable, periodic, and chaotic dynamical systems has seen an explosive growth over the past 2 decades. Further, as a consequence of the recognition that all systems are subject to noise, there have been recent attempts to understand the role of noise in altering the behaviour of the unperturbed system. Many discrete and continuous time systems with stochastic perturbations have been examined both analytically and numerically. As these systems may often be viewed as special cases of Markov processes, there is extensive applied literature concerning their stability properties [12]. Some of these stability results are couched in the language of classical Liapunov-type stability arguments [11].

The effects of continuously distributed stochastic perturbations clearly depend on the properties of the unperturbed system. Thus, the results of [8, 9] indicate that addition of noise to dynamical systems with highly irregular trajectories (Axiom A systems) will not result in an alteration of

0022-247X/89 $3.00

the statistical behaviour of the system since the invariant measure changes continuously with the noise level. However, in other cases [4, 5] the addition of noise results in evident changes in the dynamical behaviour of trajectories and may make the dynamics more regular by creating absolutely continuous invariant measures. Finally, the addition of noise to discrete time systems without statistical properties in the unperturbed state may lead either to a type of statistical periodicity or to statistical stability [13, 14]. A general answer to the question of how trajectories behave in the presence of perturbation appears to be quite difficult and involves the use of sophisticated topological methods in the definition of attracting sets [15].

This paper considers the effect of very general types of random perturbations of discrete time dynamical systems. For example, both the additive and parametric perturbations considered in [6, 7] are covered by the situation we treat. We make a mild assumption concerning an average contractive property for the system, and it is not necessary that the random perturbation have a density. Due to the possible lack of continuity of the values of the perturbation, the problem is most appropriately framed within the context of the evolution of measures or distributions. The problem is formulated precisely in Section 2.

The main difficulty in examining the stability properties of this system is related to the fact that the stationary distribution is not known a priori. Hence, the delicate monotonicity-type arguments employed in [3, 10] are not helpful. In Section 3, using a statistical-type contraction argument close to the methods of weak boundedness [11], it is proved that with perturbation the sequence of distributions is weakly convergent to a unique distribution independent of the initial distribution. Thus, we find that stochastic perturbations can induce interesting statictical properties in systems whose unperturbed dynamics have no statistical qualities.

Since these perturbed systems have a contracting property on the average they could, for example, result from the perturbation of a globally asymptotically stable system. However, they may equally well arise from the seemingly unrelated situation in which, at each time step, one randomly applies one of a set of transformations, some of which are not necessarily stable. Thus, the results of this paper are immediate generalizations of work on the reconstruction of fractal sets using iterated function systems [1, 2]. We explore this connection in Section 4, showing that the unique limiting fractal set to which iterated function systems converge is identical to the support of the unique limiting distribution considered in Section 3.

2. Formulation of the Problem

Consider a stochastically perturbed discrete time dynamical system of the form

$$x_{n+1} = S(x_n, \xi_n) \qquad \text{for} \quad n = 0, 1, ..., \tag{1}$$

where S is a given deterministic transformation defined on a subset $A \times V$ of $R^d \times R^1$ with values in A, and the ξ_n are independent 1-dimensional random vectors with values in V. The initial value x_0 is a d-dimensional random vector.

In our study of the behaviour of (1) we make the following assumptions:

(i) For every fixed $y \in V$ the function $S(x, y)$ is continuous in x and for every fixed $x \in A$ it is measurable in y. The set $A \subset R^d$ is closed and $V \subset R^1$ is Borel measurable.

(ii) The random vectors $\xi_0, \xi_1, ...$ are independent and have the same distribution; i.e.; the measure

$$\varphi(B) = \text{prob}(\xi_n \in B) \qquad \text{for} \quad B \subset V, \quad B \text{ Borelian}$$

is the same for all n.

(iii) For every n we have

$$E(|S(x, \xi_n) - S(z, \xi_n)|) < |x - z| \qquad \text{for} \quad x, z \in A, \quad x \neq z \tag{2}$$

and

$$E(|S(x, \xi_n)|^2) \leqslant \alpha |x|^2 + \beta, \qquad \text{for} \quad x \in A, \tag{3}$$

where E denotes the mathematical expectation and α and β are non-negative constants with $\alpha < 1$. The symbol $|\cdot|$ denotes an arbitrary norm in R^d which is not necessarily Euclidean.

We always assume that the initial random vector x_0 is independent of the sequence $\{\xi_n\}$.

Remark 1. Observe that in the case of A bounded, inequality (3) is automatically satisfied with $\alpha = 0$ and $\sqrt{\beta} = \sup\{|x| : x \in A\}$. Moreover from (ii) it follows that if (2) and/or (3) hold for some integer n then they are also true for every n.

Example 1. Consider the system (1) with $S(x, y) = [x/(1+x)] + y$ or

$$x_{n+1} = x_n/(1 + x_n) + \xi_n \tag{4}$$

on the half line $R^+ = [0, \infty)$. In this case $d=1$ and $A = R^+$. Assume that

$$\text{Prob}(\xi_n \geqslant 0) = 1, \qquad E(\xi_n^2) < \infty, \qquad n = 0, 1, \ldots . \tag{5}$$

Examples of possible forms for the perturbations ξ_n are $\xi_n = \{0, 1\}$ with

$$\text{Prob}(\xi_n = 0) = p, \qquad \text{Prob}(\xi_n = 1) = 1 - p, \tag{6}$$

or that the ξ_n are Poisson random variables $\xi_n = \{0, 1, 2, \ldots\}$ where

$$\text{Prob}(\xi_n = k) = \alpha^k e^{-\alpha}/k!. \tag{7}$$

However, we need not be restricted to a situation where the ξ_n have only discrete values and we could equally well, for example, assume that the ξ_n have a gamma distribution.

EXAMPLE 2. A two-dimensional analog of (4) is easily constructed, viz.,

$$\begin{aligned} x_{n+1}^1 &= x_n^2/(1 + x_n^1 + x_n^2) + \xi_n^1 \\ x_{n+1}^2 &= x_n^1/(1 + x_n^1 + x_n^2) + \xi_n^2, \end{aligned} \tag{8}$$

where A is now the first quadrant of the plane R^2.

Our goal is the study of the asymptotic behaviour of the sequence $\{x_n\}$. Since the ξ_n are random, the behaviour of x_n is uncertain even with a specified x_0. Thus, we adopt the strategy of studying the sequence of distributions

$$\mu_n(B) = \text{prob}(x_n \in B), \tag{9}$$

where B is a Borel subset of A.

The first step in this process is to find a recurrence relation that will give μ_{n+1} in terms of μ_n. By the Riesz representation theorem, a measure supported on A is uniquely defined by the value of the integral

$$\mu(f) = \int_A f(x)\, \mu(dx) \qquad \text{for} \quad f \in C_0(A),$$

where $C_0(A)$ denotes the space of all real valued continuous functions on A with compact support. Thus we would like to know the values of $\mu_{n+1}(f)$ if $\mu_n(f)$ is given. Since the vector x_n is completely determined by x_0 and $\xi_0, \ldots, \xi_{n-1}$, it is clear that x_n and ξ_n are independent. Let $f \in C_0(A)$ be given. Then the mathematical expectation of $f(S(x_n, \xi_n))$ is just

$$\int_A \int_V f(S(x, y))\, \mu_n(dx)\, \varphi(dy).$$

However, since $f(x_{n+1}) = f(S(x_n, \xi_n))$ the mathematical expectation is also given by

$$\int_A f(x)\, \mu_{n+1}(dx).$$

Equating these two expressions we immediately obtain

$$\int_A f(x)\, \mu_{n+1}(dx) = \int_A \int_V f(S(x, y))\, \mu_n(dx)\, \varphi(dy),$$

or

$$\mu_{n+1}(f) = P\mu_n(f),$$

where

$$P\mu(f) = \int_A \int_V f(S(x, y))\, \mu(dx)\, \varphi(dy). \tag{10}$$

The operator P maps the space $M(A)$ of all probabilistic measures (distributions) on A into itself. Thus, for a given initial measure μ_0, the sequence $\{P^n\mu_0\}$ describes the evolution of measures corresponding to the dynamical system (1). Alternately, using the terminology of Foias, we could say that $\{P^n\mu_0\}$ is the statistical solution of (1) corresponding to the initial distribution μ_0. In Section 3 we prove that, under the conditions stated, the sequence of distributions (9) is weakly convergent to a unique distribution that is the fixed point of the operator P defined by (10).

The utility of examining the evolution of the distributions is immediately apparent when one wishes to calculate averages of some particular quantity. For example, let $C(A)$ be the space of all continuous functions, and let $f \in C(A)$ and an initial measure μ_0 be given. Further, define $m_n(f)$ by

$$m_n(f) = \int_A f(x)\, \mu_n(dx). \tag{11}$$

By our previous comments,

$$m_n(f) = \langle f, \mu_n \rangle = \langle f, P^n\mu_0 \rangle = \langle U^n f, \mu_0 \rangle = \int_A U^n f(x)\, \mu_0(dx), \tag{12}$$

where $U\colon C(A) \to C(A)$ is the operator adjoint to P. Thus, in order to calculate $m_n(f)$ we need an expression for U. From (10),

$$P\mu(f) = \int_A \left[\int_V f(S(x, y))\, \varphi(dy) \right] \mu(dx) = \int_A Uf(x)\, \mu(dx),$$

and thus

$$Uf(x) = \int_V f(S(x, y))\, \varphi(dy). \tag{13}$$

Now take $f(x) = x^k$ so

$$M_k^n = \int_B U^n x^k \mu_0(dx)$$

is the kth-order moment after n iterations.

3. The Unique Convergence of Measures

We now turn to our main results, first noting that by use of the expression for the adjoint operator U given in (13) we may rewrite Eq. (10) in the form $P\mu(f) = \mu(U(f))$, or for general n,

$$P^n\mu(f) = \mu(U^n(f)). \tag{14}$$

Setting

$$S_1(x, y) = S(x, y) \qquad \text{and} \qquad S_{n+1}(x, y_0, ..., y_n) = S(S_n(x, y_0, ..., y_{n-1}), y_n)$$

we may write an explicit expression for Eq. (14), namely

$$U^n f(x) = \int \underset{V^n}{\cdots} \int f(S_n(x, y_0, ..., y_{n-1}))\, \varphi(dy_0) \cdots \varphi(dy_{n-1}),$$

or, introducing $Y = (y_0, ..., y_{n-1})$ and $\Phi_n = \varphi \times \overset{n}{\cdots} \times \varphi$, we have

$$P^n\mu(f) = \int_A U^n f(x)\, \mu(dx), \qquad U^n f(x) = \int_{V^n} f(S_n(x, Y))\, \Phi_n(dY). \tag{15}$$

Now denote by δ_x the measure supported at the point x, that is

$$\delta_x(B) = \begin{cases} 1 & \text{for} \quad x \in B \\ 0 & \text{for} \quad x \notin B. \end{cases}$$

If the initial point $x_0 \in A$ in (1) is fixed (nonrandom), then the sequence of random vectors $\{x_n\}$ corresponding to (1) is given by

$$x_n = S_n(x_0, \xi_0, ..., \xi_{n-1}) \tag{16}$$

and $P^n \delta_{x_0}$ is the distribution of the x_n. Our first observation concerning the behaviour of the system described by Eq. (1) is the following:

THEOREM 1. *If conditions* (i)–(iii) *hold for Eq.* (1), *then for every two points* $\bar{x}$, $\bar{\bar{x}} \in A$ *the corresponding sequences*

$$\bar{x}_n = S_n(\bar{x}, \xi_0, ..., \xi_{n-1}), \qquad \bar{\bar{x}}_n = S_n(\bar{\bar{x}}, \xi_0, ..., \xi_{n-1}),$$

satisfy

$$\lim_{n \to \infty} E(|\bar{x}_n - \bar{\bar{x}}_n|) = 0. \tag{17}$$

Proof. Consider the sequence $\{x_n\}$ given by (1) with a fixed x_0. From (3) we have

$$E(|x_{n+1}|^2) = E(|S(x_n, \xi_n)|^2) = \int_A E(|S(x, \xi_n)|^2)\, \mu_n(dx)$$
$$\leqslant \alpha \int_A |x|^2\, \mu_n(dx) + \beta = \alpha E(|x_n|^2) + \beta,$$

where $\mu_n = P^n\, \delta_{x_0}$. Consequently,

$$E(|x_n|^2) \leqslant L, \qquad L = [\beta/(1-\alpha)] + 1, \tag{18}$$

for sufficiently large n, say $n > n_0(x_0)$. Now define

$$u_n(\bar{x}, \bar{\bar{x}}) = E(|\bar{x}_n - \bar{\bar{x}}_n|) = \int_{V^n} |S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)|\, \Phi_n(dY). \tag{19}$$

It is clear that the sequence of functions $u_n\colon A \times A \to R$ is decreasing. In fact, repeated applications of the definition of S_n leads to

$$u_{n+1}(\bar{x}, \bar{\bar{x}}) = \int_{V^n} \left\{ \int_V |S(S_n(\bar{x}, Y), y_n) - S(S_n(\bar{\bar{x}}, Y), y_n)|\, \varphi(dy_n) \right\} \Phi_n(dY)$$

or

$$u_{n+1}(\bar{x}, \bar{\bar{x}}) = \int_{V^n} E(|S(S_n(\bar{x}, Y), \xi_n) - S(S_n(\bar{\bar{x}}, Y), \xi_n)|)\, \Phi_n(dY) \tag{20}$$

which, in conjunction with inequality (2), gives

$$u_{n+1}(\bar{x}, \bar{\bar{x}}) \leqslant \int_{V^n} |S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)|\, \Phi_n(dY) = u_n(\bar{x}, \bar{\bar{x}}).$$

Thus, for every $\bar{x}, \bar{\bar{x}} \in A$ the sequence of functions $u_n(\bar{x}, \bar{\bar{x}})$ is convergent. Our goal is to prove that the limiting function to which the u_n converge is identically equal to zero.

Suppose that the limiting function is not zero, and assume that for some fixed $\bar{x}, \bar{\bar{x}} \in A$

$$\sigma = \lim_{n \to \infty} u_n(\bar{x}, \bar{\bar{x}}) > 0.$$

Now, in order to use (2) and (20) effectively, define

$$A_{rq} = \{(x, z) \in A^2 \colon |x| \leqslant q,\ |z| \leqslant q,\ |x - z| \geqslant r\},$$

and

$$\omega_{rq} = \max_{A_{rq}} [E(|S(x, \xi_n) - S(z, \xi_n)|)/|x - z|]$$

for $r > 0$ and $q > 0$. Because of the continuity of $S(x, y)$ with respect to x, and inequality (3), the expectation $E(|S(x, \xi_n) - S(z, \xi_n)|)$ is a continuous function of (x, z). Thus, the maximum in the definition of ω_{rq} exists whenever the set A_{rq} is nonempty. Moreover, because of inequality (2) we have $\omega_{rq} < 1$ and evidently

$$E(|S(x, \xi_n) - S(z, \xi_n)|) \leqslant \omega_{rq} |x - z| \qquad \text{for} \quad (x, z) \in A_{rq}. \tag{21}$$

For every fixed n, r, and q we may write V^n as the union of the set

$$V^n_{rq} = \{Y \in V^n \colon (S_n(\bar{x}, Y), S_n(\bar{\bar{x}}, Y)) \in A_{rq}\},$$

and the set $W^n_{rq} = V^n \setminus V^n_{rq}$. Using this partition of V^n we may write the integral on the right hand side of Eq. (20) as the sum of two integrals. Applying inequality (21) to the integral over V^n_{rq}, and inequality (2) to the remaining integral over W^n_{rq} we obtain

$$u_{n+1}(\bar{x}, \bar{\bar{x}}) \leqslant \omega_{rq} \int_{V^n_{rq}} |S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)| \, \Phi_n(dY)$$
$$+ \int_{W^n_{rq}} |S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)| \, \Phi_n(dY),$$

or

$$u_{n+1}(\bar{x}, \bar{\bar{x}})$$
$$\leqslant u_n(\bar{x}, \bar{\bar{x}}) + (\omega_{rq} - 1) \int_{V^n_{rq}} |S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)| \, \Phi_n(dY). \tag{22}$$

Using the definition of V^n_{rq} this gives

$$u_{n+1}(\bar{x}, \bar{\bar{x}}) \leqslant u_n(\bar{x}, \bar{\bar{x}}) + r(\omega_{rq} - 1) \, \Phi_n(V^n_{rq}). \tag{23}$$

Now we are going to evaluate $\Phi_n(V_{rq}^n)$ for large n. First, using (18) with $x_0 = \bar{x}$ and a classical Chebyshev-type argument we obtain

$$\Phi_n(\bar{W}_q^n) \leqslant (L/q^2),$$

$$\int_{\bar{W}_q^n} |S_n(\bar{x}, Y)| \, \Phi_n(dY) \leqslant (L/q) \qquad \text{for} \quad n \geqslant n_0(\bar{x}),$$

where $\bar{W}_q^n = \{ Y \in V^n : |S_n(\bar{x}, Y)| > q \}$, and an analogous inequality for the set $\bar{\bar{W}}_q^n = \{ Y \in V^n : |S_n(\bar{\bar{x}}, Y)| > q \}$. Thus we have

$$\begin{aligned}
&\int_{\bar{W}_q^n} |S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)| \, \Phi_n(dY) \\
&\quad \leqslant \int_{\bar{W}_q^n} |S_n(\bar{x}, Y)| \, \Phi_n(dY) + \int_{\bar{W}_q^n} |S_n(\bar{\bar{x}}, Y)| \, \Phi_n(dY) \\
&\quad \leqslant (q/L) + \int_{\bar{W}_q^n \cap \bar{\bar{W}}_q^n} |S_n(\bar{\bar{x}}, Y)| \, \Phi_n(dY) \\
&\qquad + \int_{\bar{W}_q^n \setminus \bar{\bar{W}}_q^n} |S_n(\bar{\bar{x}}, Y)| \, \Phi_n(dY) \\
&\quad \leqslant (2q/L) + q\Phi_n(\bar{W}_q^n \setminus \bar{\bar{W}}_q^n) \leqslant (3q/L).
\end{aligned}$$

The same inequality holds for the integral of $|S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)|$ over $\bar{\bar{W}}_q^n$ and, as a consequence,

$$\begin{aligned}
&\int_{W_{rq}^n} |S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)| \, \Phi_n(dY) \\
&\qquad \leqslant (6q/L) + \int_{Z_{rq}^n} |S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)| \, \Phi_n(dY),
\end{aligned}$$

where $Z_{rq}^n = W_{rq}^n \setminus (\bar{W}_q^n \cup \bar{\bar{W}}_q^n)$. In the set Z_{rq}^n we also have $|S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)| \leqslant r$ which gives

$$\begin{aligned}
&\int_{W_{rq}^n} |S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)| \, \Phi_n(dY) \\
&\qquad \leqslant (6q/L) + r\Phi_n(Z_{rq}^n) \leqslant (6q/L) + r.
\end{aligned}$$

Fixing q and r such that $(6q/L) + r \leqslant (\sigma/2)$ and using (22) we conclude that

$$\int_{V_{rq}^n} |S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)| \, \Phi_n(dY) \geqslant u_{n+1}(\bar{x}, \bar{\bar{x}}) - (\sigma/2) \geqslant \sigma/2.$$

Since we have $|S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)| \leqslant 2q$ in the set V^n_{rq}, from the last inequality it follows that $\Phi_n(V^n_{rq}) \geqslant \sigma/4q$. From this and (23) we finally obtain

$$u_{n+1}(\bar{x}, \bar{\bar{x}}) \leqslant u_n(\bar{x}, \bar{\bar{x}}) - \delta \qquad \text{for large } n,$$

where $\delta = r\sigma(1 - \omega_{rq})/4q$ is a strictly positive number independent of n. This implies the convergence of $u_n(\bar{x}, \bar{\bar{x}}) \to -\infty$ as $n \to \infty$ which is impossible. Thus we must have $\sigma = 0$ and the proof is completed.

Remark 2. Applying the classical Dini theorem to the decreasing sequence $\{u_n\}$ we conclude that the convergence in (17) is uniform on compact subsets of A^2. Further, using inequality (3) it is easy to prove that (17) holds for every two initial random vectors $\bar{x}$, $\bar{\bar{x}}$ independent of $\{\xi_n\}$ and having finite second order moments. However, we will not use this fact in what follows.

To describe the behaviour of $P^n\mu$ precisely we require two definitions. Remember that $M(A)$ is the set of all probabilistic Borel measures on A. A measure $\mu \in M(A)$ is called *stationary* if $P\mu = \mu$. Further, a sequence $\{\mu_n\} \in M$ is called *weakly convergent* to a measure $\mu \in M$ if $\mu_n(f) \to \mu(f)$ for every $f \in C_0(A)$. With these two definitions we are ready to state our main result in:

THEOREM 2. *Assume that the maping $S: A \times V \to A$ in Eq.* (1) *and the sequence of random vectors $\{\xi_n\}$ satisfy conditions* (i)–(iii). *Then there exists a unique stationary $\mu_* \in M(A)$, and for every $\mu_0 \in M(A)$ the sequence $\{P^n\mu_0\}$ is weakly convergent to μ_*.*

Proof. The existence of a stationary measure μ_* may be proved in a standard fashion. Choose a measure $\delta_{x0} \in M(A)$ and define

$$\mu_*(f) = \operatorname{Lim} P^n \delta_{x0}(f) = \operatorname{Lim} \delta_{x0}(U^n f) \qquad \text{for} \quad f \in C_0(A), \tag{24}$$

where Lim denotes a Banach limit. Clearly $\mu_*(f)$ is a linear functional on $C_0(A)$ and thus, by the Riesz representation theorem, represents a non-negative measure. We are going to show that μ_* is normalized. In fact, from (18) and the Chebyshev inequality it follows that the measures $\mu_n = P^n \delta_{x0}$ satisfy

$$\mu_n(A_r) \geqslant 1 - (L/r^2) \qquad \text{for} \quad n \geqslant n_0(x_0),$$

where $A_r = \{x \in A : |x| \leqslant r\}$. Consequently, for every $f \in C_0(A)$ satisfying $1 \geqslant f \geqslant 1_A$ we have $1 \geqslant \mu_n(f) \geqslant 1 - L/r^2$ which, in turn, implies $1 \geqslant \mu_*(f) \geqslant 1 - L/r^2$. Since $r > 0$ is arbitrary, this proves that the measure μ_* is probabilistic. Finally, from (24) it follows that $\mu_* = \mu_*(Uf)$ or $\mu_* = P\mu_*(f)$ which shows that the measure μ_* is stationary.

Now let $\mu_0 \in M(A)$ be arbitrary and let $f \in C_0(A)$ be Lipschitzean. We have by formula (15)

$$P^n\mu_0(f) - \mu_*(f) = P^n\mu_0(f) - P^n\mu_*(f)$$
$$= \int_A U^n f(x)\, \mu_0(dx) - \int_A U^n f(x)\, \mu_*(dx).$$

Let $\varepsilon \in (0, 1)$ be an arbitrary number. Choose a compact set $A_0 \subset A$ such that $\mu_0(A \setminus A_0) \leqslant \varepsilon$ and $\mu_*(A \setminus A_0) \leqslant \varepsilon$. Define

$$\tilde{\mu}_0(B) = \mu_0(A_0 \cap B)/\mu_0(A_0), \qquad \tilde{\mu}_*(B) = \mu_*(A_0 \cap B)/\mu_*(A_0).$$

Then

$$|P^n\mu_0(f) - \mu_*(f)|$$
$$\leqslant \left| \int_{A_0} U^n f(x)\, \tilde{\mu}_0(dx) - \int_{A_0} U^n f(x)\, \tilde{\mu}_*(dx) \right| + 4\varepsilon \max |f|. \qquad (25)$$

Using the continuity of $U^n f$ we can find points $\bar{x} = \bar{x}(n)$ and $\bar{\bar{x}} = \bar{\bar{x}}(n)$ in A_0 such that

$$\left| \int_{A_0} U^n f(x)\, \tilde{\mu}_0(dx) - \int_{A_0} U^n f(x)\, \tilde{\mu}_*(dx) \right| \leqslant |U^n f(\bar{x}) - U^n f(\bar{\bar{x}})|. \qquad (26)$$

Now, using (15) and denoting the Lipschitz constant for f by L we obtain

$$|U^n f(\bar{x}) - U^n f(\bar{\bar{x}})| \leqslant L \int_{V^n} |S_n(\bar{x}, Y) - S_n(\bar{\bar{x}}, Y)|\, \Phi_n(dY)$$
$$= L u_n(\bar{x}, \bar{\bar{x}}) + 4\varepsilon, \qquad (27)$$

where the u_n are defined by formula (19). Since $\{u_n\}$ converges to zero uniformly on A_0^2, from (25), (26), and (27) it follows that $\{P^n\mu_0(f)\}$ converges to $\mu_*(f)$ for every Lipschitzean $f \in C_0(A)$. Finally, since the set of Lipschitzean functions is dense in $C_0(A)$ this implies that $\{P^n\mu_0\}$ converges weakly to μ_*, and the proof is completed.

In the proof of Theorem 2, the use of inequality (2) is evident. However, we would like to stress the role of inequality (3) in our considerations. First, observe that the existence of a stationary density depends solely on (3), since we have not used inequality (2) in the existence part of the proof. Moreover, the following example shows that (2) without (3) does not imply the convergence of $\mu_n = P^n\mu_0$ to a stationary density.

EXAMPLE 3. Consider a dynamical system on the real line ($A = V = R$) of the form

$$x_{n+1} = S(x_n) + \xi_n, \qquad (28)$$

where the ξ_n are equally distributed independent random variables such that $\xi_n \leqslant 0$ with probability one, and $S: R \to R$ is a C^1 function satisfying

$$\tfrac{1}{2} < S'(x) < 1, \qquad S(x) < x \qquad \text{for} \quad x \in R \tag{29}$$

(for example, $S(x) = (3x/4) + (\sqrt{1+x^2})/4$). It is clear that the condition $|S'(x)| < 1$ implies (2). On the other hand, from (29) it follows that $S^{-n}(c)$ converges to $+\infty$ for every $c \in R$. Thus we have

$$\begin{aligned} \mu_n((c, \infty)) &= \text{prob}(x_n > c) \leqslant \text{prob}(S^n(x_0) > c) \\ &= \text{prob}(x_0 > S^{-n}(c)) \to 0 \end{aligned} \tag{30}$$

for every $c \in R$ which demonstrates that there is no limiting distribution for the sequence $\{u_n\}$.

Finally, we simply note that from the proof of Theorem 2 it is easy to show that inequality (3) may be replaced by a weaker requirement, namely

$$E(|S(x, \xi_n)|^p) \leqslant \alpha |x|^p + \beta \qquad \text{for} \quad x \in R, \tag{31}$$

where $p > 1$, $0 \leqslant \alpha < 1$, $\beta \geqslant 0$ are constants.

4. Relation to Iterated Function Systems

In this section we consider the relation between the results of Section 4 and those of Barnsley and his co-workers [1, 2] on the limiting properties of iterated function systems. In the situations considered in [1, 2], all results were obtained for systems on an arbitrary compact metric space. To compare our results with those from iterated function systems, we assume that the Barnsley system is defined on a subset A of R^d but we do not find it necessary to assume that A is compact.

Consider a system $w_i: A \to R^d$ $(i = 1, ..., N)$ of continuous functions defined on a closed set $A \subset R^d$. For a given subset B of A set

$$w(B) = \bigcup_{i=1}^{N} w_i(B). \tag{32}$$

When B is compact then $w(B)$ is also compact. Thus w may be considered as a mapping of a space $H(A)$ of all compact and nonempty subsets of A into itself. The space $H(A)$ endowed with the Hausdorff distance

$$H(B_1, B_2) = \max_{x \in B_1} \min_{z \in B_2} |x - z| + \max_{x \in B_2} \min_{z \in B_1} |x - z|$$

becomes a metric space.

The problem that has been examined in the behaviour of iterated function systems is related to the asymptotic behaviour of $w^n(B)$. As shown in [2], this asymptotic behaviour is closely related to a stochastic process which can be roughly described as follows. Fix a probability vector $p_1, ..., p_N$ with $p_i > 0$ and $\sum p_i = 1$. Choose $x_0 \in A$ and successively define the sequence $\{x_n\}$ by choosing

$$x_{n+1} \in \{w_1(x_n), ..., w_N(x_n)\} \qquad \text{for} \quad n = 0, 1, ... \tag{33}$$

in such a way that $x_{n+1} = w_i(x_n)$ with probability p_i.

We can easily reformulate the iterated function system of [1, 2] within the framework of this paper. Assume that V is the set of all sequences $(0, ..., 1, ..., 0)$ where 1 is in the ith place, $i = 1, ..., N$. Further, consider a sequence of independent random vectors $\{\xi_n\}$ with values in V such that

$$\text{Prob}(\xi_n^i = 1) = p_i, \tag{34}$$

where ξ_n^i denotes the ith coordinate of ξ_n. Now define $S: A \times V \to A$ by setting

$$S(x, y) = \sum_{i=1}^{N} w_i(x)\, y^i, \tag{35}$$

where again y^i denotes the ith coordinate of the vector $y = (y^1, ..., y^N)$. Now,

$$x_{n+1} = S(x_n, \xi_n) \tag{36}$$

gives the required sequence of random variables.

Given a measure $\mu \in M(A)$ we will denote its support by $\operatorname{supp} \mu$. Thus $x \in \operatorname{supp} \mu$ if and only if

$$\mu(K(x, \delta)) > 0 \qquad \text{for every} \quad \delta > 0,$$

where $K(x, \delta)$ is an open ball with center x and radius δ. From the definition it follows that $\operatorname{supp} \mu$ is a closed set. Now let $\{\mu_n\}$ denote the sequence of distributions corresponding to the random variables $\{x_n\}$ described by Eqs. (35) and (36). Assume that $B = \operatorname{supp} \mu_0$ is compact. Using the continuity of w_i it is straightforward to verify that $\operatorname{supp} \mu_1 = w(B_0)$ and, by induction,

$$\operatorname{supp} \mu_n = w^n(B), \qquad \text{for} \quad n = 0, 1, 2, ... \tag{37}$$

Thus, the behaviour of $w^n(B)$ is completely described by the behaviour of μ_n. However, when the limiting set

$$B_* = \lim_{n \to \infty} w^n(B) \tag{38}$$

exists (in the sense of the Hausdorff distance on $H(A)$) it may not be equal to the support of the limiting measure μ_*. That this may be the case is easily shown by the following.

EXAMPLE 4. Let $A=[0,1]$, $N=2$, and let $w_1 \equiv 0$, $w_2 \equiv x$, and $p_1 = p_2 = \frac{1}{2}$. In this case $V = \{\{0,1\}, \{1,0\}\}$ and

$$S(x, y) = xy^2, \qquad y = (y^1, y^2). \tag{39}$$

We have

$$E(|S(x, \xi_n) - S(z, \xi_n)|) = |x - z|\; E(\xi_n^2) = \tfrac{1}{2}|x - z|.$$

Thus, according to Theorem 2 the limiting measure μ_* exists and is unique. Actually, μ_* can be easily calculated in this case since

$$U(f) = E(f(S(x, \xi_0))) = E(f(x, \xi_0^2)) = \tfrac{1}{2}f(0) + \tfrac{1}{2}f(x),$$

and generally

$$U^n(f) = (1 - 2^{-n}) f(0) + 2^{-n} f(x).$$

Therefore,

$$P^n\mu(f) = \int_{[0,1]} U^n f(x)\, \mu(dx) = (1 - 2^{-n}) f(0) + 2^{-n} \int_{[0,1]} f(x)\, \mu(dx)$$

and, as a consequence, we have

$$\lim_{n \to \infty} P^n\mu(f) = f(0).$$

Thus the limiting measure $\mu_* = \delta_0$ and $\operatorname{supp} \mu_* = \{0\}$. However, on the other hand we have

$$w(B) = w_1(B) \cup w_2(B) = \{0\} \cup B,$$

and by induction $w^n(B) = \{0\} \cup B$. Thus $w^n(B)$ converges to $\{0\} \cup B$ which differs from the support of μ_* unless $B = \{0\}$. It is obvious that in this particular example the set $\{0\}$ is the most appropriate limiting set. In fact, all but one of the sequences $\{x_n\}$ described by (33) and (34) have the form

$$\underbrace{x_0, \ldots, x_0}_{n+1}, 0, 0, , \ldots, \qquad n = 0, 1, \ldots$$

with corresponding probabilities 2^{-n}. The unique sequence $x_n = x_0$ appears with probability zero.

The critical fact in this example leading to the lack of correspondence between the limiting set B_* and the support of the measure μ_* is the dependence of $B_* = \{0\} \cup B$ on the choice of the initial set B. The contrary situation is covered by the following.

PROPOSITION 1. *Assume that the functions $w_i: A \to R^d$ $(i = 1, ..., N)$ are continuous and such that the limiting set $B_* \in H(A)$ defined by* (38) *exists for every $B \in H(A)$ and is independent of B. Further, assume that for every $\mu \in M(A)$ there exists a limiting distribution $\mu_* = \lim \mu_n$ independent of μ. Then the support of μ_* is B_*.*

Proof. Choose an arbitrary $\mu_0 \in M(A)$ with compact support B. Set $B_n = w^n(B)$. From (37) we have $B_n = \operatorname{supp} \mu_n$. The weak convergence of μ_n to μ_* implies that

$$\operatorname{supp} \mu_* \subset \lim_{n \to \infty} \operatorname{supp} \mu_n = \lim_{n \to \infty} B_n = B_*.$$

Now let B_∞ be the support of μ_*. Since B_∞ is a closed set contained in B_* it is compact. Consider $B = B_\infty$ as the initial set for $B_n = w^n(B)$ and $\mu_0 = \mu_*$ as the initial distribution. We have $\mu_n = \mu_*$ so, again by (37),

$$B_n = \operatorname{supp} \mu_n = \operatorname{supp} \mu_*,$$

which, in turn, implies that

$$B_* = \lim_{n \to \infty} B_n = \operatorname{supp} \mu_*,$$

which completes the proof.

The iterated function systems $\{w^n\}$ considered in [1, 2] are restricted to the particular case when all of the w_i are strictly contractive, i.e., they satisfy

$$|w_i(x) - w_i(z)| \leqslant \lambda |x - z| \qquad \text{for} \quad x, z \in A, \tag{40}$$

with a constant $\lambda < 1$, where A is a compact set. Thus, in a situation where the contractive property (40) holds then the assumptions of Proposition 1 are always satisfied. That this is the case may be easily seen by noting that, first, the inequality (40) implies, according to [1, 2], that the limit of B_* of (38), always exists and is independent of the initial B. Secondly, for the transformation S defined by (35) we always have

$$\begin{aligned} E(|S(x, \xi_n) - S(z, \xi_n)|) &= E\left(\left|\sum_{i=1}^{N} (w_i(x) - w_i(z))\, \xi_n^i\right|\right) \\ &\leqslant |x - z| \sum_{i=1}^{N} E(\xi_n^i) = |x - z|, \end{aligned}$$

which is a much stronger condition than (2), which is assumed in the statement and proof of Theorem 2. Finally, one must simply recall that, by Remark 1, (3) is automatically satisfied for compact sets.

Iterated function systems have many potential applications, and many examples may be found in [1, 2]. In closing we would like to add one further example which is quite unexpected.

EXAMPLE 5. Consider the baker transformation

$$T(u, v) = \begin{cases} (2u, \frac{1}{2}v), & 0 \leqslant u < \frac{1}{2}, 0 \leqslant v \leqslant 1 \\ (2u-1, \frac{1}{2}v + \frac{1}{2}), & \frac{1}{2} \leqslant u \leqslant 1, 0 \leqslant v \leqslant 1, \end{cases}$$

and observe that the first (u) coordinate of T simply evolves by the dyadic transformation $T_1(u) = 2u \pmod 1$. Set

$$(u_{n+1}, v_{n+1}) = T(u_n, v_n) \qquad \text{for} \quad (u_0, v_0) \in [0, 1] \times [0, 1].$$

Then these iterates of the baker transformation may be written in the form

$$\begin{aligned} u_{n+1} &= T_1(u_n) \\ v_{n+1} &= \tfrac{1}{2}v_n + \xi_n, \end{aligned} \tag{41}$$

where the ξ_n are given by

$$\xi_n = 1_{[1/2, 1]}(u_n).$$

From a classical result of Borel, it is known that the ξ_n defined in this way are independent random variables on the probability space $([0, 1], \mathscr{B}, \mu)$ where $\mathscr{B}$ is the sigma algebra of all Borel subsets of $[0, 1]$ and μ is the usual Borel measure.

Equation (41) may be rewritten as

$$v_{n+1} = S(v_n, \xi_n)$$

with $S(v, y) = \frac{1}{2}v + y$. It is clear that S and $\{\xi_n\}$ satisfy all of the assumptions for iterated function systems as well as the conditions of Theorem 2. Thus, by the simple expedient of taking the v projection of the Baker transformation an iterated function system is produced.

ACKNOWLEDGMENTS

We thank A. Kyle and L. Mackey for helpful discussions, and Prof. James D. Murray, Director, Centre for Mathematical Biology, Mathematics Research Institute, University of Oxford, England, for his hospitality during the time this work was completed. This research was partially supported by the Natural Sciences and Engineering Research Council (Canada) through Grant A-0091 and by the Science and Engineering Research Council (Great Britian) through Grant GR/D/13573.

References

1. M. F. Barnsley and D. Demko, Iterated function systems and the global construction of fractals, *Proc. Roy. Soc. London Ser. A* **399** (1985), 243–275.
2. M. F. Barnsley, E. Ervin, D. Hardin, and J. Lancaster, Solution of an inverse problem for fractals and other sets, *Proc. Nat. Acad. Sci. U.S.A.* **83** (1986), 1975–1977.
3. K. Baron and M. Kuczma, Iteration of random-valued functions on the unit interval, *Colloq. Math.* **37** (1977), 263–269.
4. A. Boyarsky, Randomness implies order, *J. Math. Anal. Appl.* **76** (1980), 483–497.
5. A. Boyarsky, On the significance of absolutely continuous invariant measures, *Phys. D* **11** (1984), 130–146.
6. J. P. Crutchfield and B. A. Huberman, Fluctuations and the onset of chaos, *Phys. Lett. A* **77** (1980), 407–410.
7. J. P. Crutchfield, J. D. Farmer, and B. A. Huberman, Fluctuations and simple chaotic dynamics, *Phys. Rep.* **92** (1982), 45–82.
8. Ju. I. Kifer, On the limiting behaviour of invariant measures of small random perturbations for some smooth dynamical systems, *Dokl. Akad. Nauk SSSR* **216** (1974), 339–369.
9. Ju. I. Kifer, On small perturbations of some smooth dynamical systems, *Math. USSR-Izu.* **8** (1974), 1083–1107.
10. M. Kuczma, Normalizing factors for iterates of random valued functions, *Prace Nauk. Uniw. Ślask. Katowic.* **87** (1975), 67–72.
11. H. J. Kushner, "Stochastic Stability and Control," Academic Press, New York/London, 1967.
12. H. J. Kushner, "Introduction to Stochastic Control Theory," Holt, Rinehart and Winston, New York, 1972.
13. A. Lasota and M. C. Mackey, "Probabilistic Properties of Deterministic Systems," Cambridge Univ. Press, Cambridge, 1985.
14. A. Lasota and M. C. Mackey, Noise and statistical periodicity, *Phys. D* **28** (1987), 143–154.
15. D. Ruelle, Small random perturbations of dynamical systems and the definition of attractors, *Comm. Math. Phys.* **82** (1981), 137–151.