AD-A119 195 BROWN UNIV PROVIDENCE RI LEFSCHETZ CENTER FOR DYNAMI--ETC F/G 12/1
AN INVARIANT MEASURE APPROACH TO THE CONVERGENCE OF STOCHASTIC --ETC(U)
MAR 82 H J KUSHNER, A SHWARTZ N00014-76-C-0279
UNCLASSIFIED LCDS-82-15 AFOSR-TR-82-0722 NL

END
DATE FILMED
10 82
DTIC

AFOSR-TR- 82-0722

(2)

AD A119195

Lefschetz Center for Dynamical Systems

DTIC

SEP 1 4 198

Approved for publ
distribution un

Division of Applied Mathematics   Brown University Providence RI 02912

An Invariant Measure Approach to the Convergence
of Stochastic Approximations with
State Dependent Noise

2

by

H. J. Kushner and Adam Shwartz

March 1982 LCDS #82-15

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 82-0722	2. GOVT ACCESSION NO. AD-A119 195	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AN INVARIANT MEASURE APPROACH TO THE CONVERGENCE OF STOCHASTIC APPROXIMATIONS WITH STATE DEPENDENT NOISE		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
		6. PERFORMING ORG. REPORT NUMBER LCDS #82-15
7. AUTHOR(s) Harold J. Kushner and Adam Shwartz		8. CONTRACT OR GRANT NUMBER(s) AFOSR-81-0116
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lefschetz Center for Dynamical Systems, Division of Applied Mathematics, Brown University Providence RI 02912		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE61102F; 2304/A4
11. CONTROLLING OFFICE NAME AND ADDRESS Directorate of Mathematical & Information Sciences Air Force Office of Scientific Research Bolling AFB DC 20332		12. REPORT DATE March 1982
		13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A new method is presented for quickly getting the ODE (ordinary differential equation) associated with the asymptotic properties of the stochastic approximation $X_{n+1} = X_n + a_n f(X_n, \xi_n)$ (or the projected algorithm for the constrained problem). Either $a_n \to 0$, or a_n can be constant, in which case

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ITEM #20, CONTINUED:

the analysis is on the sequence obtained when $a \to 0$.) The method basically requires that $\{X_n, \xi_{n-1}\}$ be Markov with a "Feller" transition function, but little else. The simplest result requires that if $X_n \equiv x$, the corresponding noise process $\{\xi_n(x), n \geq 0\}$ have a unique invariant measure; but the 'non-unique' case can also be treated. No mixing condition is required, nor the construction of averaged test functions, and $f(\cdot,\cdot)$ need not be continuous. A detailed analysis of the way that $\{\xi_n\}$ varies with $\{X_n\}$ is not required. For the class of sequences treated, the conditions seem easier to verify than for other methods. There are extensions to the non-Markov case. Two examples illustrate the power and ease of use of the approach. Aside from the advantages of the method in treating standard problems, it seems to be particularly useful for handling the type of iterative algorithms which arise in adaptive communication theory, where the dynamics are often discontinuous and the 'noise' is often state dependent due to the effects of feedback.

An Invariant Measure Approach to the Convergence of Stochastic Approximations with State Dependent Noise*

by

Harold J. Kushner+

Adam Shwartz++

Divisions of Applied Mathematics and Engineering
Lefschetz Center for Dynamical Systems
Brown University
Providence, Rhode Island 02912

March 1982

* This research has been supported in part by the Air Force Office of Scientific Research under contract #AFOSR 81-0116, in part by the National Science Foundation under contract #NSF-Eng. 77-12946-A04 and in part by the Office of Naval Research under contract #N00014-76-C-0279-P0004.

+ Division of Applied Mathematics and Engineering
Brown University, Providence, Rhode Island 02912

++ Division of Engineering, Brown University, Providence, Rhode Island 02912

An Invariant Measure Approach to the

Convergence of Stochastic Approximations with State Dependent Noise

by

Harold J. Kushner and Adam Shwartz

Abstract

A new method is presented for quickly getting the ~~ODE~~ (ordinary differential equation) associated with the asymptotic properties of ~~the~~ stochastic approximation ~~$X_{n+1} = X_n + a_n f(X_n, \xi_n)$~~ (or the projected algorithm for the constrained problem). Either $a_n \to 0$, or a_n can be constant, in which case the analysis is on the sequence obtained when $a \to 0$.) The method basically requires that the stochastic approximation ~~$\{X_n, \xi_{n-1}\}$~~ be Markov with a "Feller" transition function, but little else. The simplest result requires that if $X_n \equiv x$, (is equivalent to) the corresponding noise process ~~$\{\xi_n(x), n \geq 0\}$~~ have a unique invariant measure; but the 'non-unique' case can also be treated. No mixing condition is required, nor the construction of averaged test functions, and $f(\cdot,\cdot)$ need not be continuous. A detailed analysis of the way that $\{\xi_n\}$ varies with $\{X_n\}$ is not required. For the class of sequences treated, the conditions seem easier to verify than for other methods. There are extensions to the non-Markov case. Two examples illustrate the power and ease of use of the approach. Aside from the advantages of the method in treating standard problems, it seems to be particularly useful for handling the type of iterative algorithms which arise in adaptive communication theory, where the dynamics are often discontinuous and the 'noise' is often state dependent due to the effects of feedback.

I. Introduction

We consider stochastic approximations of the form

(1.1) $$X_{n+1} = X_n + a_n f(X_n, \xi_n),$$

where $f(\cdot,\cdot)$ might be discontinuous, and the evolution of $\{\xi_n\}$ depends on $\{X_n\}$ in the sense that, in general,

$$P\{\xi_{n+1} \in A | \xi_i, i \leq n\} \neq P\{\xi_{n+1} \in A | X_i, \xi_i, i \leq n\}.$$

We also treat the following "projected" version of (1.1). Let G be a bounded set of the form $G = \{x : q_i(x) \leq 0, i = 1, \dots, s\}$, where $q_i(\cdot)$ are continuously differentiable, and G is the closure of its interior. Let $\pi_G(y)$ denote any closest point in G to y. Then the projected algorithm is

(1.2) $$X_{n+1} = \pi_G(X_n + a_n f(X_n, \xi_n)).$$

Several so-called ordinary differential equations (ODE) methods for proving convergence of $\{X_n\}$ have been developed in recent years. ([1] to [4], and [5], a more polished form of [4], with weaker conditions). The aim of these methods is to get an ODE, which we write symbolically (for (1.1)) as

(1.3) $$\dot{x} = E^x f(x,\xi) \equiv \int f(x,\xi) P^x(d\xi)$$

where (loosely speaking) $P^x(\cdot)$ is the stationary distribution of the sequence $\{\xi_n\}$, when $X_n \equiv x$. The idea is that $\{X_n\}$ in (1.1) varies much more slowly (for large n) than $\{\xi_n\}$ does and that some sort of averaging method or law of large numbers can be used to show that the asymptotic properties of $\{X_n\}$ are the same as those of (1.3), with a proper definition of $P^x(\cdot)$.

The methods in [1] to [3] are very useful, but are often difficult to apply when the noise is state dependent, in the sense that the conditions are often either hard to verify or do not hold in many important cases of interest. Reference [4], [5] presented an "averaging method" which works quite well for such problems, although one would like to avoid the work associated with constructing the "averaged test functions", and verifying the conditions on them. The results in [4], [5] were for w. p. 1. convergence and also proved stability and similar properties for $\{X_n\}$ sequences which were not artificially bounded. But generally, past methods required what is often a difficult analysis of the way $\{\xi_n\}$ depends on $\{X_n\}$.

In this paper, the essential assumption for the validity of (1.3) is that $\{\xi_n\}$ depends on $\{X_n\}$ in such a way that if $X_n \equiv x$, a constant, then the corresponding $\{\xi_n\}$ process possesses a unique stationary measure. Such an assumption, either implicitly or explicitly, was used in much past work on the 'state-dependent' noise case. If the stationary measures are not unique, then a very similar result (2.9) holds. The conditions required here are generally weaker and much easier to check and are useful even when the noise does not depend on the state. As amply shown by the examples, the method is easy to use. The techniques used are new for the class of problems treated.

We concentrate on the case $a_n \to 0$. The same proofs work (even more easily) when $a_n \equiv a$, a constant. Then, we get that the limit (as $a \to 0$) of $x^a(\cdot)$ satisfies (1.3) or (2.12) (in the constrained case), where $x^a(\cdot)$

is the piecewise linear function with values X_n at na. The approach is advantageous for treating many standard problems because the conditions are relatively easy to verify. They are particularly useful for treating the type of algorithms which appear in adaptive communication theory, where the dynamics are often discontinuous and the noise is often state dependent owing to the role of feedback. In such cases, one normally has $a_n \equiv a$.

In Section 2, we discuss the case where $\{X_n, \xi_{n-1}, n \geq 1\}$ is a Markov process, or where the 'state-noise' pair can be 'Markovianized'. This is the case which is most fully understood and easiest to use. A class of non-Markov processes is dealt with in Section 3, and in Section 4 we illustrate the power and ease of use of the method via two examples.

2. Limit Theorems with $\{X_n, \xi_{n-1}\}$ Markov.

In this section, we are concerned with the Markov case. In very many applications the system is either Markovian or the actual physical noise can be Markovianized, perhaps leading to an abstract valued process. Below, it is assumed that $\{X_n\}$ is either tight or lies in a compact set. This is not a very serious restriction, since practical algorithms tend to use various truncation devices. In any case, the use of the projection method (1.2) guarantees the compactness when X_n lies in a Euclidean space.

In Theorems 1 and 2, we allow X_n to take values in a compact metric space, and ξ_n in a complete separable metric space. The reason for this is that it fits certain 'abstract' applications where the metric is defined by a weak topology, and which will be published separately. Also, it facilitates the treatment of non-Markovian problems by 'Markovianizing' in an abstract space.

Some assumptions. Assumptions (A2.4,6) and the first part of (A2.1) will be weakened later, as will the explicit form given for $f(\cdot,\cdot)$.

A2.1. $\{X_n, \xi_{n-1}, n \geq 0\}$ is a Markov process with a (possibly non-homogeneous) transition function $P(x,\xi,n,\ell,A) = P\{(X_{n+\ell},\xi_{n+\ell-1}) \in A | X_n = x, \xi_{n-1} = \xi\}$. The X_n takes values in a compact subset H of a metric and linear topological space, and ξ_n takes values in S, a complete separable metric space. Both metrics are assumed to be invariant.

A2.2. $\{\xi_n\}$ is tight in S.

A2.3 For each Borel $B \subset S$, define the one step 'transition function' $P_x(\xi,1,B) = P\{\xi_n \in B | \xi_{n-1} = \xi, X_n = x\}$, and suppose that it does not depend on n. Let $P_x(\xi,1,\cdot)$ be weakly† continuous in (x,ξ).

For each x, we now define a Markov process $\{\xi_n(x), n \geq 0\}$ via the transition function $P_x(\xi,\ell,\cdot)$, where $P_x(\xi,\ell,B) = \int P_x(\xi,\ell-k,d\xi')P_x(\xi',k,B)$ is defined recursively.

A2.4. For each $x \in H$, let $\{\xi_n(x), n \geq 0\}$ have a unique invariant measure $P^x(\cdot)$, and let $\{P^x(\cdot), x \in H\}$ be tight††.

A2.5. $\sum_n |a_{n+1} - a_n| < \infty$, $0 < a_n \to 0$, $\sum_n a_n = \infty$

A2.6, $f(\cdot)$ is bounded

A2.7. There is an integer $c \geq 0$ such that $\int P_x(\xi,c+1,d\xi')f(x,\xi')$ is continuous in (x,ξ). It equals $\lim_j \int P(x,\xi,j-c,c,dx'\, d\xi')P_{x'}(\xi',1,d\xi'')f(x',\xi'') = \lim_j E[f(X_j,\xi_j)|X_{j-c}=x,\xi_{j-c-1}=\xi]$, where the limit is uniform on compact (x,ξ) sets.

† I.e., $\int P_x(\xi,1,d\xi')g(\xi)$ is continuous in (x,ξ) if $g(\cdot)$ is bounded and continuous.

†† Normally, the tightness holds when (A2.2) holds, so the condition is not restrictive.

Remark on (A2.7). If $f(\cdot,\cdot)$ is continuous, then (A2.3) implies that we can take $c = 0$. If $c = 0$, then the second sentence of (A2.7) is redundant. Even if $f(\cdot,\cdot)$ is not continuous $c = 0$ is often enough to get the required smoothing. See, for example, the applications in Section 4. Even if $c > 0$ is needed, the second sentence of (A2.7) does not seem to be particularly restrictive, since $|X_j - X_{j-c}| \to 0$ as $j \to \infty$ implies that the measure in the $\lim_j$ expression is essentially $P_x(\xi, c+1, \cdot)$ for large j. The assumption is stated as it is for technical reasons. In all applications that we are aware of now, if the first sentence of (A2.7) holds, so does the second sentence.

Before introducing the next assumption, some additional notation is required. Set $t_n = \sum_0^{n-1} a_i$, and $m(t) = \max\{n : t_n \leq t\}$, for $t \geq 0$. Thus $m(t_n) = n$. Let $0 < \delta_n \to 0$ as $n \to \infty$ such that $\limsup_n \{a_j : j \geq n\}/\delta_n = 0$. For each n choose a sequence $\{m(\ell,n), \ell=1,\ldots\}$ where $m(n,1) = n$, $m(n,\ell+1) > m(n,\ell)$, and such that $\sum_{m(n,\ell)}^{m(n,\ell+1)-1} a_j = \delta_n$, modulo an 'end' value of a_j. Thus $(t_{m(n,\ell+1)} - t_{m(n,\ell)})/\delta_n \to 1$ as $n \to \infty$, uniformly in ℓ. For notational convenience we henceforth suppress the n in $m(\ell,n)$ and write simply $m(n,\ell) = m_\ell$. Let $I_K(\xi)$ denote the indicator of the set where $\xi \in K$. For each ω, ℓ, n define the measure on the Borel sets of S:

$$Q(\omega,\ell,n,\cdot) = \frac{1}{\delta_n} \sum_{m_\ell}^{m_{\ell+1}-1} a_j \, P\{\xi_{j-1} \in \cdot \,|\, X_{m_\ell}, \xi_{m_\ell - 1}\},$$

Define $Q_K(\omega,\ell,n,\cdot) = Q(\omega,\ell,n,\cdot) I_K(\xi_{m_\ell - 1})$. Thus, if $\xi_{m_\ell - 1}(\omega) \notin K$, the measure is the zero measure. If S is not compact, then another assumption is needed. First, we state it (A2.8b) and then discuss it. Either (A2.8a) or (A2.8b) will be used. (A2.8b) always holds for $N(K) = 1$ if S is compact.

A2.8a. Either S is compact or $\{\xi_n\}$ is mutually independent.

or

A2.8b. For each compact K, there is an integer $N(K) < \infty$ such that for each T the set

$\{P(X_{m_\ell}(\omega), \xi_{m_\ell - 1}(\omega), m_\ell, j-m_\ell, \omega) I_{\{\xi_{m_\ell - 1}(\omega) \in K\}};$ <u>all compact</u> K, <u>all</u> n, ℓ, j, ω,

<u>such that</u> $j \geq m_\ell \geq n,\ j-m_\ell \geq N(K),\ \ t_j - t_n \leq T\}$

<u>is tight</u>.

Despite its seemingly complicated structure, (A2.8b) is quite natural and is often easy to verify. See, for example, the application in Section 4b, and the example below. It is motivated by the following consideration. If S is not compact, it is possible that

$$(*) \qquad \{P\{\xi_{j-1} \in \cdot \mid X_{m_\ell}(\omega), \xi_{m_\ell - 1}(\omega)\},\ \ j \geq m_\ell,\ \omega\}$$

is not tight. Suppose for example that $\{\xi_n\}$ is a stationary scalar valued Gauss Markov process, not depending on $\{X_n\}$, and whose correlation function $\rho(\cdot)$ tends to zero. Then the set (*) is not tight, since arbitrarily large initial conditions $\xi_{m_\ell - 1}(\omega)$ are allowed. But, if the $\xi_{m_\ell - 1}(\omega)$ were all confined to a bounded set, then (*) would be tight. As K increases, and $\xi_{m_\ell - 1}(\omega) \in K$, it might take longer for the 'future' $\xi_j (j > m_\ell)$ to 'settle down'. This is why we allow $N(K)$ steps for this 'settling down', where $N(K)$ increases with K. In this example, if $K = \{\xi : |\xi| \leq k\}$, then any $N(K)$ satisfying $\rho(N(K)) \cdot k \leq$ constant is satisfactory.

We now take some notation from [2]. Let $x^0(\cdot)$ denote the piecewise linear interpolation of the function with value X_n at t_n. Define the shifted function $x^n(\cdot)$ by $x^n(t) = x^0(t+t_n)$, $t \geq 0$. Thus $x^n(0) = X_n$, and the asymptotic properties (as $t \to \infty$) of any limit (as $n \to \infty$) of $\{x^n(\cdot)\}$ yields the asymptotic behavior of $\{X_n\}$. The convergence of $x^n(\cdot)$ to a limit $x(\cdot)$ is in the sense of weak convergence of a sequence of probability measures. We give the differential equations which $x(\cdot)$ satisfies. Using this differential equation and the properties of weak convergence, one can analyze the asymptotic behavior of X_n as in [2]. See that reference for details. Here, we concentrate or proving representations for the differential equations.

Theorem 1. Under (A2.1) to (A2.8), $\{x^n(\cdot)\}$ is tight and any weak limit $x(\cdot)$ satisfies

$$\dot{x} = E^x f(x,\xi) = \int f(x,\xi)P^x(d\xi) \quad \text{w.p.1}, \tag{2.1}$$

where $x(0) \in H$. The right hand side of (2.1) is continuous in x.

Proof. For notational simplicity, we do the proof only for the case where H is a subset of a Euclidean space E^r. The details in the general case are quite similar. What is actually proved in the general case is that for each bounded real valued $g(\cdot)$ whose first two Frechet derivatives are continuous, $g(x(t)) - g(x(0)) = \int_0^t g_x(x(u))oE^{x(u)}f(x(u),\xi)du$ w.p.1.

The continuity is a consequence of the tightness and uniqueness (A2.4).

Now, by the tightness of $\{\xi_n\}$, we can choose $\delta_n \to 0$ and non-decreasing compact K_α such that

$$P\{\xi_{\nu_\alpha} \in K_\alpha\} \to 1 \text{ for any sequence } \{\nu_\alpha\}, \text{ and}$$

$$\frac{1}{\delta_n} \sum_{m_\ell}^{m_\ell+N(K_n)} a_i \to 0, \text{ uniformly in } \ell \text{ (where } m_\ell \geq n\text{), as } n \to \infty,$$

and

$$\frac{1}{\delta_n} \sum_{j \geq n} |a_{j+1}-a_j| \to 0 \text{ as } n \to \infty.$$

Define the piecewise constant function $\tilde{f}_n(\omega,t)$ by

$$\tilde{f}_n(\omega,t) = \frac{1}{\delta_n} \sum_{m_\ell}^{m_{\ell+1}-1} a_j E_{m_\ell} f(X_j,\xi_j) I_{K_n}(\xi_{m_\ell-1}) \text{ on } [t_{m_\ell}-t_n, t_{m_{\ell+1}}-t_n),$$

and set $\tilde{F}_n(\omega,t) = \int_0^t \tilde{f}_n(\omega,s)ds$. (To see why time is 'centered' about t_n, when working with the shifted function $x^n(\cdot)$, recall that $t_m - t_n = \sum_{j=n}^{m-1} a_j$ are the break points of $x^n(\cdot)$ and that $x^n(0) = x^0(t_n) = X_n$.) By (A2.6), $\{x^n(\cdot), \tilde{F}_n(\cdot), n \geq 0\}$ is tight in $C^{2r}[0,\infty)$. Henceforth, we work with a weakly convergent subsequence (called subsequence 1), also indexed by n, and with limit $(x(\cdot),\tilde{F}(\cdot))$, or with a subsequence of it. We use Skorokhod imbedding ([7], Theorem 3.1.1) wherever convenient, and with no notational change. Thus, we can assume where convenient, that w.p.1. $(x^n(\cdot),\tilde{F}_n(\cdot))$ converges to $(x(\cdot),\tilde{F}(\cdot))$ uniformly on bounded time intervals. Suppose that

$$\tilde{F}(t) = \int_0^t E^{x(u)} f(x(u),\xi)du \tag{2.2}$$

and that for arbitrary k, t, s and $s_1 < s_2 \ldots < s_k < t < t + s$ and bounded and continuous $h(\cdot)$,

$$Eh(x(s_j),\tilde{F}(s_j),j \leq k)[x(t+s)-x(t)-\tilde{F}(t+s)-\tilde{F}(t))] = 0. \tag{2.3}$$

Then $M(t) = x(t) - x(0) - \tilde{F}(t)$ is a continuous martingale with $M(0) = 0$. Since the quadratic variation of $M(\cdot)$ is zero (as can readily be shown), $M(t) \equiv 0$ w.p.1, and (2.1) holds. So, we only need to prove (2.2) and (2.3).

For smooth $h(\cdot)$,

$$Eh(x^n(s_j),\tilde{F}_n(s_j),j \leq k)[x^n(t+s)-x^n(t) - \sum_{m(t_n+t)}^{m(t_n+t+s)-1} a_j f(X_j,\xi_j)] \equiv \varepsilon_n \tag{2.4}$$

$$Eh(x^n(s_j),\tilde{F}_n(s_j),j \leq k)[x^n(t+s)-x^n(t) - \int_t^{t+s} \tilde{f}_n(s)ds] \equiv \varepsilon_n', \tag{2.5}$$

where ε_n and ε_n' go to zero as $n \to \infty$.

We now prove

$$\tilde{f}_n(s) \to E^{x(s)} f(x(s),\xi) \text{ \underline{in probability for each} } s. \tag{2.6}$$

The limit (2.6) implies that $\tilde{F}_n(\cdot)$ converges in measure to the right side of (2.2). This and the weak convergence and (2.5) yield (2.3) with $\tilde{F}(\cdot)$ defined by (2.2). So, only (2.6) needs to be proved.

<u>Fix</u> s <u>and</u> $\{m_\ell\}$ <u>such that</u>[†] $s \in [t_{m_\ell}-t_n, t_{m_\ell+1}-t_n)$. Let N_0 denote the null set on which $(x^n(\cdot),\tilde{F}_n(\cdot))$ does not converge uniformly to $(x(\cdot),\tilde{F}(\cdot))$ on bounded intervals (under the Skorokhod imbedding). It is enough to show that (for each s) each subsequence of subsequence 1 contains a further subsequence for which the limit in (2.6) holds in probability. Select a subsequence of subsequence 1, indexed also by n but called subsequence 2, such that[††] $P\{\xi_{m_\ell-1} \in K_n,$ all large $n\} = 1$. Let N(s) denote the exceptional null set. <u>Fix</u> $\omega \notin N_0 \cup N(s)$, and extract a weakly convergent subsequence (a subsequence of subsequence 2) of the set of measures (tight by (A2.8) and the properties of δ_n) $Q = \{Q_{K_n}(\omega,\ell,n,\cdot): n \in$ subsequence 2, s fixed as above$\}$, with limit $\bar{P}_\omega(\cdot)$. The limits in (2.7) below are on this subsequence. Let $g(\cdot)$ be bounded and continuous and set $G(x,\xi) = \int P_x(\xi,1,d\xi')g(\xi')$. Then by (A2.3,5,8)

$$\begin{aligned}
&\lim_n \sum_{m_\ell}^{m_{\ell+1}-1} \frac{a_j}{\delta_n} \int P(X_{m_\ell},\xi_{m_\ell-1},m_\ell,j-m_\ell,d\xi)g(\xi)I_{K_n}(\xi_{m_\ell-1}) \\
&= \lim_n \sum_{m_\ell}^{m_{\ell+1}-1} \frac{a_j}{\delta_n} \int P(X_{m_\ell},\xi_{m_\ell-1},m_\ell,j-m_\ell-1,d\xi',dx')P_{x'}(\xi'1,d\xi)g(\xi)I_{K_n}(\xi_{m_\ell-1}) \\
(2.7)\quad &= \lim_n \int Q_{K_n}(\omega,\ell,n,d\xi')G(X_{m_\ell},\xi') \\
&= \int \bar{P}_\omega(d\xi')G(x(s),\xi') = \int \bar{P}_\omega(d\xi')P_x(\xi',1,d\xi)g(\xi).
\end{aligned}$$

---

† I.e., for each n, choose $m_\ell = m(\ell,n)$ such that s is in the indicated interval. Keep in mind that ℓ depends on n, and that we suppress the n-dependence of $m(\ell,n)$ in the notation.

†† By the tightness of $\{\xi_j\}$, we can always choose such a subsequence.

Similarly, the limit in the first line is $\int \overline{P}_\omega(d\xi)g(\xi)$. In going from the second to the third line of (2.7), we used the facts that $G(\cdot,\xi')$ is continuous, uniformly on compact ξ', that $\sup_{m_{\ell+1} \geq j \geq m_\ell} |X_j - X_{m_\ell}| \to 0$ as $n \to \infty$, that $X_{m_\ell} \to x(s)$ and the tightness of the set Q.

Due to the arbitrariness of $g(\cdot)$, and the uniqueness (A2.4) and to the equality of the last line of (2.7) with $\int \overline{P}_\omega(d\xi)g(\xi)$, we have $\overline{P}_\omega(\cdot) = P^{x(s)}(\cdot)$. Again, by the uniqueness, the limit does not depend on the chosen subsequence of Q. Thus, if $\omega \notin N_0 \cup N(s)$, $Q_{K_n}(\omega,\ell,n,\cdot) \to P^{x(s)}(\cdot)$ weakly as $n \to \infty$, where n now indexes the second chosen 'subsequence 2' of the theorem.

Using (A2.7) we now have (limits are on the 'subsequence 2')

$$\begin{aligned}
&\lim_n \sum_{m_\ell}^{m_{\ell+1}-1} \frac{a_j}{\delta_n} \int P(X_{m_\ell}, \xi_{m_\ell-1}, m_\ell, j-m_\ell, d\xi', dx) P_x(\xi',1,d\xi) f(x,\xi) I_{K_n}(\xi_{m_\ell-1}) \\
&= \lim_n \sum_{m_\ell}^{m_{\ell+1}-1} \frac{a_j}{\delta_n} \int P(X_{m_\ell},\xi_{m_\ell-1},m_\ell,j-m_\ell-c,d\xi',dx') P(x',\xi',j-c,c,d\xi,dx) P_x(\xi,1,d\xi'') f(x,\xi'') I_{K_n}(\xi_{m_\ell-1}) \\
&= \lim_n \sum_{m_\ell}^{m_{\ell+1}-1} \frac{a_j}{\delta_n} \int P(X_{m_\ell},\xi_{m_\ell-1},m_\ell,j-m_\ell-c, d\xi'\, dx) P_x(\xi',c+1,d\xi) f(x,\xi) I_{K_n}(\xi_{m_\ell-1}) \\
&= \lim_n \sum_{m_\ell}^{m_{\ell+1}-1} \frac{a_j}{\delta_n} \int P(X_{m_\ell},\xi_{m_\ell-1},m_\ell,j-m_\ell-c,d\xi') I_{K_n}(\xi_{m_\ell-1}) [P_{X_{m_\ell}}(\xi',c+1,d\xi) f(X_{m_\ell},\xi)] \qquad (2.8) \\
&= \lim_n \int Q_{K_n}(\omega,\ell,n,d\xi') P_{x(s)}(\xi',c+1,d\xi) f(x(s),\xi) \\
&= \int \overline{P}_\omega(d\xi') P_{x(s)}(\xi',c+1,d\xi) f(x(s),\xi) \\
&= \int P^{x(s)}(d\xi) f(x(s),\xi).
\end{aligned}$$

In going from the 3rd to the 4th and then to the 5th line we used the continuity in (x,ξ) of $\int P_x(\xi',c+1,d\xi)f(x',\xi)$ and the facts concerning convergence cited below (2.7). In going from the next to last step to the last step, we used the fact that $\overline{P}_\omega(\cdot) = P^{x(s)}(\cdot)$ is an invariant measure for the transition function $P_{x(s)}(\xi,j,\cdot)$ for each $x(s)$. The equality of the first and last lines of (2.8) for $\omega \notin N_0 \cup N(s)$ yields the desired result (2.6) for 'subsequence 2'. But, since each subsequence of subsequence 1 contains a further subsequence satisfying the requirements of subsequence 2 (but with perhaps a different $N(s)$) (2.6) holds for subsequence 1 also. Furthermore, the limits for all possible subsequence 1's differ only in the initial condition $x(0)$. Q.E.D.

Extension. In many problems of interest (see Section 4), the algorithm (1.1) takes the form $X_{n+1} = X_n + a_n f_n(\omega)$, where

$$E[f_n(\omega)|X_i,\xi_{i-1},i \le n] \equiv F_n(X_n,\xi_{n-1}),$$

and $F_n(x,\xi) \to F(x,\xi)$, a continuous function, uniformly in (x,ξ) on compact sets. Then Theorem 1 still holds. This extension is useful when $f_n(\cdot)$ depends on variables other than (X_n,ξ_n); for example, it might depend on a 'choice' or 'logical' variable Z_n, where $P(Z_n=1|X_i,\xi_{i-1},i \le n) = q(X_n,\xi_{n-1})$, for some continuous function $q(\cdot)$.

Non unique $P^x(\cdot)$. A very similar result to Theorem 1 can be obtained when the uniqueness in (A2.4) is dropped. Let $\mathcal{P}^x = \{P^x_\alpha(\cdot), \alpha \in$ some set $A(x)\}$ denote the set of invariant measures for the transition function $P_x(\xi,j,\cdot)$. Assume that $\mathcal{P} = \{\mathcal{P}^x, x \in$ any compact set$\}$ is tight. For each x,

$\mathcal{P}^x$ is convex and weakly compact. Define the set

$$C(x) = \{y : y = \int P_\alpha^x(d\xi) f(x,\xi), \quad \alpha \in A(x)\}.$$

Then $C(x)$ is closed and convex. The sets $\mathcal{P}^x$ and $C(x)$ are upper semi-continuous in x in the Hausdorff topology (with the metrized weak topology on the space of distributions and the metric topology on H). This is a consequence of the fact that under the tightness of $\mathcal{P}$, if $x_n \to x$ and $P^{x_n}(\cdot) \in \mathcal{P}^{x_n}$, then $\{P^{x_n}(\cdot)\}$ is tight, and all weak limits are in $\mathcal{P}^x$ by (A2.3).

Theorem 2. Assume (A2.1)-(A2.8), with (A2.4) altered as above, then $\{x^n(\cdot)\}$ is tight and any weak limit $x(\cdot)$ satisfies

(2.9) $$\dot{x} \in C(x) \quad \text{for almost all } \omega, t.$$

Remarks on the proof. The proof is essentially the same as that of Theorem 1, and we only remark on a couple of points. By the argument of Theorem 1, if $\omega \notin N_0 \cup N(s)$ is fixed and n indexes a weakly convergent subsequence of the set of measures Q defined above (2.7), then we must have

(2.10) $$\lim_n \sum_{m_\ell}^{m_{\ell+1}-1} \frac{a_j}{\delta_n} \int P(X_{m_\ell}, \xi_{m_\ell - 1}, m_\ell, j-m_\ell, dx', d\xi') f(x',\xi') I_{K_n}(\xi_{m_\ell - 1})$$

$$= \lim_n \tilde{f}_n(s) = \int f(x(s),\xi) P_\alpha^{x(s)}(d\xi) \in C(x(s)),$$

for some $\alpha \in A(x(s))$, perhaps depending on ω and s and on the selected subsequence. Under the weak convergence for the selected subsequence and the Skorokhod imbedding, $\tilde{F}_n(\cdot) \to \tilde{F}(\cdot)$ (which is absolutely continuous) uniformly on bounded time intervals w.p.1, but $\tilde{f}_n(s)$ does not necessarily converge in probability to $\tilde{f}(s) = \dot{\tilde{F}}(s)$ as it did in Theorem 1. But, for each fixed

$\omega \notin N_0$ and each $T < \infty$, $\tilde{f}_n(\cdot)$ (considered as a function on $[0,T]$) converges (along any 'subsequence 1') to $\tilde{f}(\cdot)$ weakly when these functions are considered as elements of $L_1[0,T]$. Thus for each $\omega \notin N_0$ there are $\{\beta_{ni}, i \le n\}$ such that $0 \le \beta_{ni}$, $\sum_{i=1}^{n} \beta_{ni} = 1$, $\beta_{ni} \to 0$ as $n \to \infty$ for each i, and $\sum_{i=1}^{n} \beta_{ni}\tilde{f}_i(\cdot) \to \tilde{f}(\cdot)$ in the norm of $L_1[0,T]$. This convergence, *together with the limit* (2.10), the convexity and closure of $C(x)$ and the upper semi-continuity cited above Theorem 2 imply that $\tilde{f}(s) \in C(x(s))$ for almost all (ω,s).

The projection algorithm (1.2).

Recall the definition of π_G from Section 1. Let $\bar{\pi}(h(\cdot))$ denote the (not necessarily unique) projection of the vector field $h(\cdot)$ onto G; i.e.,

$$\bar{\pi}(h(x)) = \text{set of limits } \lim_{\Delta \downarrow 0} [\pi_G(x+\Delta h(x))-x]/\Delta. \tag{2.11}$$

Theorem 3. Assume (1.2) and the conditions above it instead of (1.1), and assume (A2.1) to (A2.8), except that X_n takes values in a Euclidean space. Then $\{x^n(\cdot)\}$ is tight and if $x(\cdot)$ is the limit of a weakly convergent subsequence, $x(\cdot)$ satisfies the 'projected' equation

$$\dot{x} = \bar{\pi}(E^x f(x,\xi)) \quad \text{for almost all } \omega,t. \tag{2.12}$$

Recall the extension of Theorem 1 to the algorithm $X_{n+1} = X_n + a_n f_n(\omega)$, cited after Theorem 1. If $\pi_G(X_n+a_n f_n(\omega))$ is used, then Theorem 3 holds with $\dot{x} = \bar{\pi}(E^x F(x,\xi))$.

Remarks on the proof. The proof is quite similar to that of Theorem 1, and only a few remarks will be made. Use the partition of [2, eqn.(5.3.4)] to write (1.2) in the form

$$X_{n+1} = X_n + a_n f(X_n,\xi_n) + a_n d_n, \tag{2.13}$$

where $a_n d_n = \pi_G(X_n + a_n f(X_n,\xi_n)) - (X_n + a_n f(X_n,\xi_n))$. In [2], $a_n d_n$ is termed τ_n. Using the notation of Theorem 1, define the piecewise constant function $\tilde{d}_n(\omega,t)$ by

$$\tilde{d}_n(\omega,t) = \frac{1}{\delta_n} \sum_{m_\ell}^{m_{\ell+1}-1} a_j d_j \qquad \text{on } [t_{m_\ell} - t_n, t_{m_{\ell+1}} - t_n),$$

and set $\tilde{D}_n(\omega,t) = \int_0^t \tilde{d}_n(\omega,s)ds$. Then, as in Theorem 1, $\{x^n(\cdot),\tilde{F}_n(\cdot),\tilde{D}_n(\cdot)\}$ is tight. Extract a convergent subsequence with limit $(x(\cdot),\tilde{F}(\cdot),\tilde{D}(\cdot))$. Both $\tilde{F}(\cdot)$ and $\tilde{D}(\cdot)$ are absolutely continuous and $\tilde{F}(\cdot)$ satisfies (2.2). Write $\tilde{f}(\cdot) = \dot{\tilde{F}}(\cdot)$ and define $\tilde{d}(\cdot)$ by

$$\tilde{D}(t) = \int_0^t \tilde{d}(s)ds.$$

Write

$$\tilde{f}(t) = \bar{\pi}(\tilde{f}(t)) + \hat{f}(t), \quad \hat{f}(t) = \tilde{f}(t) - \pi(\tilde{f}(t)),$$

the $\hat{f}(\cdot)$ term being a 'projection error'. By the method of Theorem 1,

$$x(t) - x(0) - \tilde{F}(t) - \tilde{D}(t) = 0 \quad \text{w.p.1.} \tag{2.14}$$

$$\dot{x} = E^x f(x,\xi) + \tilde{d} = \tilde{f} + \tilde{d},$$

$$\dot{x}(t) = \bar{\pi}(\tilde{f}(t)) + \hat{f}(t) + \tilde{d}(t), \quad \text{w.p.1.}$$

Using the ideas of [2, Section 5.3], it can be shown that $-\int_0^t \hat{f}(s)ds = \tilde{D}(t)$. This and (2.14) imply (2.12). We omit the rest of the details. We note only that the proof of the last equality uses the facts that if $X_{n+1} \in \partial G$, then d_n is in the cone $-K(X_{n+1})$, and that if $x(t) \in \partial G$, then $\hat{f}(t)$ in the cone $K(x(t))$, where

$$K(x) = \{y: y = \sum_{i:q_i(x)=0} \lambda_i q_{i,x}(x), \text{ for some set of } \lambda_j \geq 0\}.$$

Unbounded $f(\cdot)$.

We will use

(A2.9) There are a $K < \infty$ and a positive valued function $d(\cdot)$ such that $|f(x,\xi)| \leq K_1(1+d(\xi))$, and x takes values in the Euclidean space R^r. For some $\alpha > 0$, $\sup_j E|d(\xi_j)|^{1+\alpha} < \infty$.

Theorem 4. Under (A2.9), and the tightness of $\{X_n\}$, both $\{x^n(\cdot)\}$ and $\{\tilde{F}_n(\cdot)\}$ are tight in $C^r[0,\infty)$.

Proof. Both $x^n(\cdot)$ and $\tilde{F}_n(\cdot)$ are sums of terms of the types $a_j f(X_j,\xi_j)$ and $a_j E_{m_\ell} f(X_j,\xi_j)$, for $j \geq m_\ell$, respectively. These are bounded by $a_j K_1(1+d(\xi_j))$ and $a_j K_1(1+E_{m_\ell} d(\xi_j))$, respectively. But by (A2.9), both $\{d(\xi_j)\}$ and $\{E_{m_\ell} d(\xi_j);\ \ell,j: j \geq m_\ell\}$ are uniformly integrable, which implies the theorem.

Q.E.D.

Given the tightness, the only further impediment to the result of Theorem 1 for the unbounded $f(\cdot,\cdot)$ case, concerns the meaning of the integrals in (2.8). A truncation and limit argument seems the most natural. We simply take the following natural approach.

Suppose that there is a sequence $\{f_L(x,\xi), L = 1,2,\ldots,\}$ each member of which satisfies (A2.6,7), where c doesn't depend on L, and such that $E^x f_L(x,\xi) \to E^x f(x,\xi)$ uniformly on any compact set, as $L \to \infty$. Let (see (A2.9)) $|f(x,\xi)| \le K(1+d(\xi))$, and let $f_L(x,\xi) = f(x,\xi)$ when $d(\xi) \le L$. For each $T < \infty$, let $E \sum_{j=n}^{m(t_n+T)} a_j d(\xi_j)\, I\{d(\xi_j) \ge L\} \to 0$ as $L \to \infty$, uniformly in $n \le m(t_n+T)$. Then under (A2.1,2,3,4,5,8), the conclusion of Theorems 1 and 3 hold. The condition of the next to last sentence is guaranteed by (A2.9) and also implies the tightness.

3. The Non-Markov Case.

The ideas of the last section can be extended to some interesting non-Markov systems where, loosely speaking, if $X_n \equiv x$ (a constant) for all $-\infty < n < \infty$, then $\{\xi_n\}$ is stationary and has certain mixing properties. We next state some assumptions, which are modifications of some in Section 2. Then a general convergence theorem is proved. Lastly, it will be shown that the assumptions hold in many cases of interest.

In particular, in Theorem 8 we verify (A3.2,3) when $\{\xi_n\}$ is not state dependent and satisfies a type of ϕ-mixing condition. This case is of interest, since the non-Markov noise and discontinuous dynamics case is usually hard and occurs frequently. In this 'non-state dependent' case, the measure $P_x(\xi,1,\cdot) = P(\xi,1,\cdot)$ below would not depend on x, and would equal the stationary conditional distribution $P\{\xi_1 \in \cdot | \xi_0 = \xi\}$ provided that this stationary measure is weakly continuous in ξ.

A3.1. $X_n \in H$, a compact metric space, and $f(\cdot)$ is bounded.

The limits in (A3.2) and (A3.3) are in the sense of probability.

A3.2. For each x, there is a transition function $P_x(\xi,1,\cdot)$ which is weakly continuous in (x,ξ) and such that for each bounded and continuous $g(\cdot)$, with $G(x,\xi) = \int P_x(\xi,1,d\xi')g(\xi')$,

$$\lim_{\substack{j-n\to\infty \\ n\to\infty}} [\int P(d\xi_j | X_j,\xi_{j-1};X_u,\xi_{u-1}, u \le n)g(\xi_j) - G(X_j,\xi_{j-1})] = 0.$$

A3.3. Define $F(x,\xi) = \int f(x,\xi')P_x(\xi,1,d\xi')$. Then $F(\cdot,\cdot)$ is continuous and

$$\lim_{\substack{j-n\to\infty \\ n\to\infty}} [\int P(d\xi_j | X_j,\xi_{j-1};X_u,\xi_{u-1},u \le n)f(X_j,\xi_j) - F(X_j,\xi_{j-1})] = 0.$$

A3.4. For the Markov process with transition function $P_x(\xi,j,\cdot)$ (which is obtained recursively from $P_x(\xi,1,\cdot)$ as above (A2.4)), there is a unique invariant measure $P^x(\cdot)$. The set S is compact. (Hence, $\{P^x(\cdot), x \in H\}$ is tight.)

We define the measure $Q(\omega,\ell,n,\cdot)$ similarly to that in Section 2: i.e., by

$$\int Q(\omega,\ell,n,d\xi)g(\xi) = \frac{1}{\delta_n} \sum_{j=m_\ell}^{m_{\ell+1}-1} a_j \int P(d\xi_j | X_u,\xi_{u-1},u \le n)g(\xi_j),$$

where $g(\cdot)$ is an arbitrary bounded measurable function. (Here, we use ξ_j in the sum $Q(\cdot)$; in Theorem 1, ξ_{j-1} was used. The choice is unimportant and is due to notational convenience.)

Theorem 5. Assume (A3.1 to 4) and (A2.5,6). Then $\{x^n(\cdot)\}$ is tight and the limit $x(\cdot)$ of any weakly convergent subsequence satisfies (2.1). If (1.2) is used in lieu of (1.1) and the conditions above (1.2) hold, with X_n in some Euclidean space, then $x^n(\cdot)$ is tight and the limit of any weakly convergent subsequence satisfies (2.12).

Proof. The proof is close to that of Theorem 1 and we use the same terminology, but with the measure $Q(\omega,\ell,n,\cdot)$ defined as above and δ_n satisfying the conditions above (A2.8), and in the proof of Theorem 1. Since S is compact the truncation factor I_K used in Theorem 1 is not required. In the proof, we take $X_n \in R^r$, Euclidean r-space. The details in general are very similar to the details in this special case. Clearly $\{x^n(\cdot),\tilde{F}_n(\cdot)\}$ is tight in $C^{2r}[0,\infty)$. Extract a weakly convergent subsequence (also indexed by n and called subsequence 1) with limit $x(\cdot),\tilde{F}(\cdot)$. We work with this sequence or subsequences of it, henceforth. By the Skorokhod imbedding, there is a null set N_0 such that the limit can be taken to be uniform on bounded intervals, for $\omega \notin N_0$.

Let $g(\cdot)$ be bounded and continuous.

By (A3.2), in getting the limit in probability (as $n \to \infty$) of an expression of the form

$$\sum_{m_\ell}^{m_{\ell+1}-1} \frac{a_j}{\delta_n} \int P(d\xi_j\,,\, dX_j|X_u,\xi_{u-1},u \le m_\ell)g(\xi_j)$$

$$(3.1)\qquad = \sum_{m_\ell}^{m_{\ell+1}-1} \frac{a_j}{\delta_n}\int P(d\xi_{j-1},dX_j|X_u,\xi_{u-1},u \le m_\ell)P(d\xi_j|X_j,\xi_{j-1},X_u,\xi_{u-1},u \le m_\ell)g(\xi_j)\quad .$$

we can substitute $G(X_j,\xi_{j-1})$ for $\int P(d\xi_j|X_j,\xi_{j-1},X_u,\xi_{u-1},u \le m_\ell)g(\xi_j)$,

when $m_{\ell+1} > j \ge m_\ell$. Fix s. For each n, fix $m_\ell = m(\ell,n)$ such that $s \in [t_{m_\ell}-t_n, t_{m_{\ell+1}}-t_n)$.

Under the Skorokhod imbedding $E|G(X_j,\xi_{j-1}) - G(x(s),\xi_{j-1})| \to 0$. Thus, by the last three sentences

$$E\left|\int Q(\omega,\ell,n,d\xi)[g(\xi) - G(x(s),\xi)]\right| \overset{n}{\to} 0.$$

Choose a subsequence (called subsequence 2) for which

$$\int Q(\omega,\ell,n,d\xi)[g(\xi) - G(x(s),\xi)] \to 0 \text{ w.p.1}$$

for a countable dense set of bounded and continuous $g(\cdot)$, hence for all bounded and continuous $g(\cdot)$. Denote the exceptional ω-set by $\tilde{N}(s)$. For fixed $\omega \notin N_0 \cup \tilde{N}(s)$, choose a further subsequence (termed subsequence 3) for which $Q(\omega,\ell,n,\cdot)$ converges weakly to some measure $\bar{P}_\omega(\cdot)$.

Then

$$\int \bar{P}_\omega(d\xi)g(\xi) = \int \bar{P}_\omega(d\xi)G(x(s),\xi) = \int \bar{P}_\omega(d\xi)P_{x(s)}(\xi,1,d\xi')g(\xi').$$

Thus, by uniqueness $\bar{P}_\omega(\cdot)=P^{x(s)}(\cdot)$, a result which does not depend on the particular chosen subsequence 3. Hence $Q(\omega,\ell,n,\cdot) \to P^{x(s)}(\cdot)$ weakly along subsequence 2, for almost all ω.

Using this last result and (A3.3), and a factorization similar to the one used in (2.8), we get that

$$\sum_{m_\ell}^{m_{\ell+1}-1} \frac{a_j}{\delta_n} \int P(d\xi_j,dX_j|X_u,\xi_{u-1},u \le m_\ell)f(X_j,\xi_j) \tag{3.2}$$

$$\to \int P^{x(s)}(d\xi)f(x(s),\xi)$$

in probability as $n \to \infty$ along subsequence 2. Hence (3.2) holds in probability as $n \to \infty$ along subsequence 1.

We omit the details for the projection algorithm. These use an adaptation of the above method which is analogous to the modification of the proof of Theorem 1 which is used in Theorem 3. Q.E.D.

Remark. Let $X_{n+1} = X_n + a_n f_n(\omega)$ replace (1.1), and suppose that there is a continuous $F(\cdot)$ such that

$$E[f_j(\omega) | X_j, \xi_{j-1}; X_u, \xi_{u-1}, u \le n] - F(X_j, \xi_{j-1}) \overset{P}{\to} 0$$

as $j - n \to \infty$ and $n \to \infty$. Then the Theorem continues to hold.

We now examine (A3.2,3), under the following ϕ-mixing condition, where the noise does not depend on the state.

A3.5. Let S be compact. Let $\mathcal{F}_j$, $\mathcal{F}_0^m$ and $\mathcal{F}_m^\infty$ denote the σ-algebras which measure ξ_{j-1}, $\{\xi_{j-1}, j \le m\}$ and $\{\xi_{j-1}, j \ge m\}$, resp. For any $A \in \mathcal{F}_0^m$, $B \in \mathcal{F}_{n+m}^\infty$, $|P(AB) - P(A)P(B)| \le \phi_n P(A)$ and $\le \phi_n P(B)$, uniformly in m, where $0 \le \phi_n \to 0$.

The following result is well known.

Lemma 6. Let $A_{n+m} \in \mathcal{F}_{n+m}^\infty$, $A_m \in \mathcal{F}_0^m$, and assume (A3.5). Then as $n \to \infty$

$$|P\{A_{n+m} | \mathcal{F}_0^m\} - P\{A_{n+m}\}| \overset{P}{\to} 0$$

$$|P\{A_m | \mathcal{F}_{n+m}^\infty\} - P\{A_m\}| \overset{P}{\to} 0$$

$$E|P\{A_m | \mathcal{F}_{n+m}^\infty\} - P\{A_m\}| I_{A_{n+m}} \le \phi_n P\{A_{n+m}\} \text{ and } \le \phi_n P\{A_m\}$$

$$E|P\{A_{n+m} | \mathcal{F}_0^m\} - P\{A_{m+n}\}| I_{A_m} \le \phi_n P\{A_{n+m}\} \text{ and } \le \phi_n P\{A_m\} .$$

Let $|g_j| \le 1$ *with* g_j *being* $\mathcal{F}_j$ *measurable. Then as* $n \to \infty$

$$E[g_{n+m}|\mathcal{F}_0^m] - Eg_{n+m} \xrightarrow{P} 0$$

$$E[g_m|\mathcal{F}_{n+m}^\infty] - Eg_m \xrightarrow{P} 0,$$

where the convergences are uniform in $\{g_j\}$, $\{A_m, A_{m+n}\}$, *and* m.

Theorem 7. Assume (A3.5). *Let* g_j *be* $\mathcal{F}_j$ *measurable with* $|g_j| \le 1$, *and let* $j > n$. *Then* $E[g_j|\mathcal{F}_0^n \cup \mathcal{F}_{j-1}] - E[g_j|\ \mathcal{F}_{j-1}] \xrightarrow{P} 0$ *as* $j - n \to \infty$, *uniformly in* j, *and* $\{g_j\}$.

Proof. Let $G_{j-1} \in \mathcal{F}_{j-1}$ and $G_{0,n} \in \mathcal{F}_0^n$. The $v_{n,j}^i$ below are uniformly bounded and $E|v_{n,j}^i|I_{G_{j-1}} \le 2\phi_{j-n}P\{G_{j-1}\}$, by Lemma 6. We have

$$\int_{G_{j-1}\cap G_{0,n}} E[g_j|\mathcal{F}_0^n \cup \mathcal{F}_{j-1}]dP = \int_{G_{j-1}} g_j I_{G_{0,n}} dP$$

$$= \int_{G_{j-1}} E[g_j I_{G_{0,n}}|\mathcal{F}_{j-1}]dP$$

$$= \int_{G_{j-1}} E\{g_j E[I_{G_{0,n}}|\mathcal{F}_{j-1}^j]|\mathcal{F}_{j-1}\}dP$$

$$= \int_{G_{j-1}} E[g_j|\ \mathcal{F}_{j-1}](P\{G_{0,n}\} + v_{n,j}^1)dP \tag{3.3}$$

$$= \int_{G_{j-1}} E[g_j|\ \mathcal{F}_{j-1}](I_{G_{0,n}} + v_{n,j}^2)dP$$

$$= \int_{G_{j-1}\cap G_{0,n}} E[g_j|\mathcal{F}_{j-1}]dP + \int_{G_{j-1}} v_{n,j}^3\ dP.$$

Now, suppose that the theorem is false. Then there is a sequence of sets $H_{j,n} \in \mathscr{F}_0^n \cup \mathscr{F}_{j-1}$ and an $\varepsilon > 0$ such that for some sequence $\{g_j\}$ and $j - n \to \infty$,

$$(3.4) \qquad \int_{H_{j,n}} \{E[g_j | \mathscr{F}_0^n \cup \mathscr{F}_{j-1}] - E[g_j | \mathscr{F}_{j-1}]\} dP \geq \varepsilon$$

(and/or $\leq -\varepsilon$, we use (3.4) only for simplicity). For each $\delta > 0$, there are sets $G_{j-1}^i \in \mathscr{F}_{j-1}$ and $G_{0,n}^i \in \mathscr{F}_0^n$ with $\{G_{j-1}^i, i=1,2\ldots\}$ disjoint and $P\{H_{j,n}^\sigma \Delta H_{j,n}\} \leq \delta$, where $H_{j,n}^\sigma = \bigcup_i [G_{j-1}^i \cap G_{0,n}^i]$. Now, re-do the calculation (3.3) with G_{j-1} and $G_{0,n}$ superscripted by i, and the integrals summed over i. For small enough $\delta > 0$, this yields a contradiction to (3.4), since $\sum_i E|v_{n,j}^3| I_{G_{j-1}^i} \to 0$ as $j - n \to \infty$. Q.E.D.

<u>Theorem 8.</u> Assume (A3.1,5). Let $\{\xi_j\}$ not depend on $\{X_j\}$; i.e., for all j

$$P\{d\xi_j | \xi_{u-1}, X_u, u \leq j\} = P\{d\xi_j | \xi_{u-1}, u \leq j\}.$$

<u>Suppose that there is a measure</u> $P(\xi,1,\cdot)$ <u>on Borel sets of</u> S <u>such that</u> $P(\cdot,1,B)$ <u>is measurable for each Borel</u> $B \subset S$. <u>For each bounded, real valued and continuous</u> $g(\cdot)$, <u>let</u>

$$(3.5) \qquad \begin{aligned} &\int g(\xi_j) P(d\xi_j | \xi_{j-1} = \xi) \to \int g(\xi') P(\xi,1,d\xi') = G(\xi) \quad \underline{\text{for all}} \ \xi, \\ &\int f(x,\xi_j) P(d\xi_j | \xi_{j-1} = \xi) \to \int f(x,\xi') P(\xi,1,d\xi') = F(x,\xi) \quad \underline{\text{for each}} \ x \ \underline{\text{and}} \ \xi, \end{aligned}$$

<u>where</u> $F(\cdot,\cdot)$ <u>and</u> $G(\cdot)$ <u>are continuous.</u> <u>Then</u> (A3.2,3) <u>hold</u>.

The proof follows from Theorem 7, by letting g_j be either $g(\xi_j)$ or $f(x,\xi_j)$.

4. Examples

4a. Application to a Routing Problem.

To illustrate the power of the method, we consider the automata routing example described in [8, Section 3]. Calls arrive at a transmitting or switching terminal at random, at discrete time instants $n = 0,1,2,\ldots,$ with P{one call arrives at n^{th} instant} $= \mu$, $\mu \in (0,1)$, P{ > 1 call arrives at n^{th} instant} $= 0$. From the terminal, there are two possible routings to the destination, route 1 and route 2; the i^{th} route has N_i independent lines and can thus handle up to N_i calls simultaneously. Let $[n, n+1)$ denote the n^{th} interval of time. The duration of each call is a random variable with a geometric distribution: P{call completed in the (n+1)st interval| uncompleted at end of n^{th} interval, route i used} $= \lambda_i, \lambda_i \in (0,1)$. The members of the double sequence of the interarrival times and call durations are mutually independent. In [8], the "gain" per step was a constant, and a detailed study was made of the rate of convergence. Here, we do a stochastic approximation version; i.e., $a_n \to 0$. But the case where $a_n \equiv a > 0$ is handled in the same way. Let $\{y_n\}$ denote a sequence of random variables with values in $[0,1]$. To get an unambiguous formulation, suppose that calls terminating in the n^{th} interval actually terminate at $n + \frac{1}{2}$, and arrivals and route assignments are at the instants $0,1,2,\ldots$. Define $\xi_n = (\xi_n^1, \xi_n^2)$ = route occupancy process (called X_n^ε in [8]), where ξ_n^i = number of lines of route i occupied at time n^+. If a call arrives at instant $n + 1$, the automaton "flips a coin", choosing route 1 with probability y_n and route 2 with probability $(1-y_n)$. If all lines of the chosen route i are occupied at instant $(n+1)^-$, then the call is switched to route $j (j \neq i)$. If all lines of route j are also occupied at instant $(n+1)^-$, then the call is rejected, and disappears from the system. The model can be generalized considerably, both in the number of lines and switching nodes, and in the input and call length statistics. Let J_{in} denote the indicator of the event {call arrives at $n+1$, is assigned first to route i and is accepted by route i}. The algorithm is (4.1), where $0 < \alpha < \beta < 1$ are truncation points, and $y_0 \in (\alpha,\beta)$. The bar $|_\alpha^\beta$ denotes truncation.

$$y_{n+1} = [y_n + a_n(1-y_n)J_{1n} - a_n y_n J_{2n}]\,|_\alpha^\beta. \tag{4.1}$$

Here, $P\{\xi_{n+1} = \xi' | y_n = y, \xi_n = \xi\}$ is a continuous function of (y,ξ). The Markov chain is $\{y_n,\xi_n\}$ not $\{X_n,\xi_{n-1}\}$. For each fixed $y \in [\alpha,\beta]$, $\{\xi_n(y), n \geq 0\}$ has a unique invariant measure $P^y(\cdot)$, and $E[J_{in}|y_\ell,\xi_\ell,\ell \leq n] = F_i(y_n,\xi_n)$, where $F_i(\cdot,\cdot)$ is a continuous function of y for each (discrete) ξ. Define $y^n(\cdot)$ as $x^n(\cdot)$ was defined. By Theorem 1 or 3 and the extension cited after the Theorem statement we immediately get the correct ODE (which must be satisfied by all the weak limits of $\{y^n(\cdot)\}$)

$$\dot{y} = [(1-y)E^y J_{1n} - yE^y J_{2n}] \quad \text{for } y \in (\alpha,\beta), \tag{4.2}$$
$$y(\cdot) \text{ stops on first hitting } \alpha \text{ or } \beta.$$

Simple! No analysis of rates of convergence of n-step transition functions, etc. is required. Also, no analysis of the x-dependence of the $\{\xi_n\}$ or $\{\xi_n(x)\}$ is required. The model upon which the analysis is [8] was based appeared in [9].

4b. <u>An adaptive quantizer</u>. Efficient quantization of signals in telecommunications systems is of considerable current interest (e.g., of voice signals in telephone systems). Let the signal process $z(\cdot)$ be sampled at instants $n\Delta$, $n = 0,1,\ldots,$ and let the samples $\{z(n\Delta)\}$ be quantized and then transmitted. Adaptive quantizers have been studied as a means to more efficient quantization the quantization scale for 'large' signals, should be different from that for 'small' signals An adaptive quantizer studied in [10,11] takes roughly the following form. We use $a_n \equiv \varepsilon$, a constant. Let $0 = \psi_0 < \psi_1 < \cdots < \psi_{L-1} < \psi_L = \infty$, $0 = \eta_1 < \eta_2 < \cdots < \eta_L$, where the ψ_i, η_i are real numbers. For a scaling parameter $y > 0$, define the quantization function $q(\cdot)$. For $z(n\Delta) > 0$, set $q(z(n\Delta)) = y\eta_i$ if $z(n\Delta) \in [y\psi_{i-1}, y\psi_i)$ and set $q(-z) = -q(z)$. The parameter y should vary with the signal power. To get the adaptive quantizer of concern, fix real numbers $0 < M_1^\varepsilon < M_2^\varepsilon < \cdots < M_L^\varepsilon < \infty$, where $M_1^\varepsilon < 1$, $M_L^\varepsilon > 1$, and set $\beta \in (0,1]$. Let $0 < y_\ell < y_u < \infty$. Then we adapt the scale y according to

$$y_{n+1}^\varepsilon = (y_n^\varepsilon)^\beta B_n^\varepsilon \Big|_{y_\ell}^{y_u}, \quad \text{where } B_n^\varepsilon = M_i^\varepsilon \text{ if } |z(n\Delta)| \in [y_n^\varepsilon \psi_{i-1}, y_n^\varepsilon \psi_i). \tag{4.3}$$

We do an asymptotic analysis of the sequence $y^\varepsilon(\cdot)$, defined as the piecewise linear interpolation of the function with values y_n^ε at time $n\varepsilon$. Let $y_0^\varepsilon = y_0 \in [y_\ell, y_u]$.

Now define $\ell_1 < \ell_2 < \cdots < \ell_L$, $\ell_1 < 0$, $\ell_L > 0$, and $\alpha > 0$ such that $\varepsilon\alpha < 1$. Then set $M_i^\varepsilon = (1+\varepsilon\ell_i)$, $\beta = 1 - \varepsilon\alpha$. Then using $y^{1-\varepsilon\alpha} = y[1-\varepsilon\alpha \log y] + O(\varepsilon^2)$, and $(1+\varepsilon b_n^2) = B_n^\varepsilon$,

$$\begin{aligned} y_{n+1}^\varepsilon &= [y_n^\varepsilon(1+\varepsilon b_n^\varepsilon) - \varepsilon\alpha\, y_n^\varepsilon \log y_n^\varepsilon + O(\varepsilon^2)]\Big|_{y_\ell}^{y_u} \\ &= [y_n^\varepsilon + \varepsilon F(y_n^\varepsilon, z(n\Delta)) + O(\varepsilon^2)]\Big|_{y_\ell}^{y_u}. \end{aligned} \tag{4.4}$$

Assume further that $Z(\cdot)$ is a stationary (finite order) Gauss Markov process with $\text{Cov}\, Z(t) > 0$ and let $z(t) = h'Z(t)$, for some vector $h \neq 0$. In this example, the noise does not depend on the state and so the analysis is quite simple, even though $z(\cdot)$ is not a bounded process. Define $EF(y,z(0)) = \bar{F}(y)$. Then $\bar{F}(y)$ has a unique zero $\bar{y}$ on $(0,\infty)$, and $\bar{F}(y)$ is positive for $y < \bar{y}$ and negative for $y > \bar{y}$ [8, Section 7]. In [8, Sections 7 to 9], there is a detailed investigation of the limit of $[y_n^\varepsilon - \bar{y}]/\sqrt{\varepsilon}$. Here, we are only concerned with the simpler question of the limit of $y^\varepsilon(\cdot)$. For some $c \geq 0$, $E_{Z(0)}F(y, z(n\Delta+\Delta+c\Delta))$ is continuous in $Z(0), y$ and tends to $\overline{F}(y)$ in the mean, uniformly in $y \in [y_\ell, y_u]$, as $c \to \infty$. This fact and the method of proof of Theorem 1 or of Theorem 3 and the extension cited after the theorems implies immediately that the weak limit of $\{y^\varepsilon(\cdot)\}$ satisfies (4.5).

$$\dot{y} = \overline{F}(y), \quad y(0) = y_0 \quad \text{if} \quad \bar{y} \in [y_\ell, y_u], \tag{4.5}$$

and if $\bar{y} \notin [y_\ell, y_u]$, $y(\cdot)$ stops on first hitting y_ℓ or y_u.

REFERENCES

[1] L. Ljung, "Analysis of recursive stochastic algorithms", IEEE Trans. on Automatic Control, AC-22, 1977, pp.551-575.

[2] H.J. Kushner, D.S. Clark, Stochastic Approximation Methods for Constrained and Unconstrained Systems, Applied Math. Science Series #26, Springer, Berlin, 1978.

[3] A. Benveniste, M. Goursat, G. Ruget, "Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms, with applications to data communication", IEEE Trans. on Automatic Control, AC-25, 1980, pp.1042-1057.

[4] H.J. Kushner, "Stochastic approximation with discontinuous dynamics and state dependent noise; w.p.1 and weak convergence", Proceedings of the Conference on Decision and Control, Albuquerque, N.M. 1980.

[5] H.J. Kushner, "On averaging method for stochastic approximations with discontinuous dynamics and state dependent noise", to appear in an Academic Press book edited by J. Rustagi in honor of H. Chernoff.

[6] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.

[7] A.V. Skorokhod, "Limit Theorems for stochastic processes", Theory of Probability and Applications, 1, 1956, pp.262-290.

[8] H.J. Kushner, Hai Huang, "Averaging methods for the asymptotic analysis of learning the adaptive systems with small adjustment rate", SIAM J. on Control and Optimization, 19, 1981, pp.635-650.

[9] K.S. Narendra, M.A.L. Thathachar, "On the behavior of a learning automaton in a changing environment with application to telephone traffic routing", preprint, Yale University, Department of Engineering, 1979.

[10] D.J. Goodman, A. Gersho, "Theory of an adaptive quantizer", IEEE Trans. on Comm., COM-22, 1974, pp.1037-1045.

[11] D. Mitra, "A generalized adaptive quantization system with a new reconstruction method for noisy transmission", IEEE Trans. Comm., COM-27, 1979, pp.1681-1689.