Stochastic Approximation with Discontinuous Dynamics and State Dependent Noise: w.p. 1 and Weak Convergence*

HAROLD J. KUSHNER

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

Stochastic approximations of the form $X_{n+1} = X_n + a_n h(X_n, \xi_n)$ are treated where $h(\cdot, \cdot)$ might not be continuous and the noise sequence $\{\xi_n\}$ might depend on $\{X_n\}$. An 'averaging' and an 'ordinary differential equation' method are combined to get w.p.1 convergence for both the above algorithm and for the case where the iterates are projected back onto a bounded set $G$ if they ever leave it. Two examples are developed, the first being an automata problem where the dynamics are not smooth and the noise is state dependent, and the second a Robbins-Monro process with observation averaging (which causes the noise to be state dependent). Each example is typical of a larger class. The results hold if $|a_n|$ is a sequence of random variables, $a_n$ being dependent on $\{X_i, i \leq n\}$. If $a_n \to 0$ but $\Sigma a_n < \infty$ does not hold, then weak convergence results are obtained.

1. INTRODUCTION

References [1, 2] present a collection of fairly general methods for proving w.p.1 and weak convergence results for stochastic approximations of the type

$$X_{n+1} = X_n + a_n h(X_n, \xi_n), \quad X_n \in \mathbb{R}^r, \text{ Euclidean } r\text{-space}, \quad (1.1)$$

where $\{\xi_n\}$ is a sequence of random variables and $0 < a_n \to 0, \Sigma a_n = \infty$. Also, several stochastic approximation schemes for sequential monte carlo function minimization or equation solving under equality and inequality constraints were dealt with. One, among others, is the projection method. Let $q_1, \ldots, q_m$ denote continuously differentiable functions, define $G = \{x: q_i(x) \leq 0, i = 1, \ldots, m\}$, then the algorithm is

$$X_{n+1} = \pi_G[X_n + a_n h(X_n, \xi_n)] \quad (1.2)$$

* This research has been supported in part under contract AFOSR-76-3063C of the Air Force Office of Scientific Research, in part by National Science Foundation under Grant NSF-Eng 77-12946, and in part by the Office of Naval Research under Grant 00014-76-C-0279-P0004.
where \( \pi_{\alpha}(y) \) denotes the closest point on \( G \) to \( y \). Both weak convergence and w.p.1 results were proved for this and several other "constrained" algorithms.

If \( h(x, \xi) \) is not additive in \( \xi \), then the methods in [1] (and also in [3], which deals with related algorithms, at least for the unconstrained case) require that \( h(\cdot, \cdot) \) be continuous. In many applications, \( h(\cdot, \cdot) \) is not continuous (e.g., \( h(\cdot, \cdot) \) might be an indicator function). Here, we combine some of the basic ideas from [1] together with the averaging methods of [4, 5] to develop an alternative method which is more convenient when \( h(\cdot, \cdot) \) is not smooth, and which is often quite advantageous if \( \{\xi_n\} \) is state dependent. We rely on the assumption that even if \( h(\cdot, \cdot) \) is not smooth, expectations or conditional expectations of the types \( Eh(x, \xi_n), E[h(x, \xi_n)|\xi_{n-1}, \xi_{n-2}, \ldots, 1] \) are smooth functions of \( x \). This situation occurs in many examples. Reference [6] also makes such an assumption for non-smooth \( h(\cdot, \cdot) \), but deals with \( a_n = a > 0 \), and a finite time interval \([n: an \leq T]\).

In Sections 2, 3, respectively, we treat the cases (1.1), (1.2), respectively, and where \( \{\xi_n\} \) is bounded and not state dependent. Section 4 deals with the case of state dependent \( \{\xi_n\} \) and the "unbounded" noise case is briefly discussed. The convergence is w.p.1 in all cases of Sections 2 through 6. Two interesting classes of examples appear in Sections 5 and 6. Weak convergence results are given in Section 7. Reference [8] contains slightly more general results.

2. The Algorithm (1.1)

Assumptions. \( E_n \) denotes expectation conditioned on \( \{\xi_j, j < n\} \). \( K \) denotes a constant whose value might change from usage to usage and \( \delta X_n \) denotes \( X_{n+1} - X_n \). Let \( \mathcal{C}_0^2 \) denote the space of \( R \)-valued functions on \( R' \) with compact support and whose second partial derivatives are continuous.

(A1) \( \Sigma a_n < \infty, \quad \Sigma a_n = \infty, \quad \{a_{n+1}/a_n\} \) is bounded, \( h(\cdot, \cdot) \) is measurable and \( h(x, \cdot) \) is bounded uniformly on bounded \( x \)-sets. \( \{\xi_n\} \) is uniformly bounded.

(A2) There is a twice continuously differentiable Liapunov function \( 0 \leq V(x) \) such that \( V_{x}(\cdot) \) is bounded, \( V(x) \to \infty \) as \( |x| \to \infty \) and for some \( \varepsilon_0 > 0 \) and compact set \( Q_0 \) of the form \( \{x: V(x) \leq \lambda_0\} \), \( V_{x}(x) \) \( \bar{V}(x) < -\varepsilon_0 \) for \( x \in Q_0 \), where \( \bar{V}(\cdot) \) is defined in (A3).

(A3) There is a continuously differentiable function \( \bar{h}(\cdot) \) and a null set \( N_0 \) such that for each \( n \) and \( x \) and \( \omega \in N_0 \), the function defined by

\[
V_n(x, n) = \sum_{j=n}^{\infty} a_j V_j(x) E_n[h(x, \xi_j) - \bar{h}(x)],
\]
is bounded by $K a_n (1 + | V'_x(x) \bar{h}(x) |)$ where the convergence for $V_0(x, n)$ and for all infinite sums of the sequel is in the sense $\lim \sum_{\infty}^n a_i \left| \right|$ for each $x$, and where the sequence of partial sums is bounded uniformly on compact $x$-sets. The bound holds for $V(\cdot)$ replaced by an arbitrary $f(\cdot) \in \mathcal{F}_a^2$.

(A4) $E_n | h(x, \xi_n) |^2 \leq K (1 + | V'_x(x) \bar{h}(x) |)$

(A5) $| V'_x(x) \bar{h}(x) | \leq K (1 + V(x))$

(A6) Let $[ \frac{\cdot}{\frac{\cdot}{\infty}} ]$ denote the gradient here. Then

$$\left| \sum_{j=0}^{\infty} a_j \left[ V'_x(x) E_{n+1}(h(x, \xi_j) - \bar{h}(x)) \right] \right| \leq K a_n (1 + | V'_x(x) \bar{h}(x) |)^{1/2}$$

The bound holds for $f(\cdot) \in \mathcal{F}_a^2$ replacing $V(\cdot)$.

(A7) For $0 \leq s \leq 1$,

$$E_n \left| V'_x(x + s a_n h(x, \xi_n)) \bar{h}(x + s a_n h(x, \xi_n)) \right| \leq K (1 + | V'_x(x) \bar{h}(x) |).$$

The examples show that the assumptions are often not restrictive.

Let $X^0(\cdot)$ denote the continuous piecewise linear function which equals $X_0$ on $[-\infty, 0]$, $X_n, n \geq 0$, at $t_n = \sum_{i=0}^{n-1} a_i$ and in each $(t_n, t_{n+1})$ is a linear interpolation of $X_n$ and $X_{n+1}$. Define $X^0(\cdot)$ by $X^0(t) = X^0(t_0 + t)$. Note that $X^0(0) = X^0(t_0) = X_n$, and define $m(t) = \max \{n: t_n \leq t\}$ for $t \geq 0$ and $m(t) = 0$ for $t < 0$.

**Theorem 1.** Assume (A1)–(A7). Then $\{X_n\}$ is bounded w.p.1. If $V'_x(x) \bar{h}(x) \leq 0$ for all $x$, then $X_n \rightarrow \{x: V'_x(x) \bar{h}(x) = 0\}$ w.p.1. In general, $\{X_n\}$ converges w.p.1 to the largest bounded invariant set of

$$\dot{x} = \bar{h}(x). \quad (2.1)$$

If $x_0 = x(t)$ is an asymptotically stable solution of (2.1) (in the sense of Liapunov) with domain of attraction $DA(x_0)$, and if $X_n \in$ compact $A \subset DA(x_0)$ infinitely often, then (except for $\omega$ in a null set) $X_n \rightarrow x_0$ as $n \rightarrow \infty$.

**Proof.** We have

$$E_n V(X_{n+1}) - V(X_n) = a_n V'_x(X_n) E_n h(X_n, \xi_n) \quad (2.2)$$

$$+ a_n^2 \int_0^1 E_n h'(X_n, \xi_n) V_{xx}(X_n + s \delta X_n) h(X_n, \xi_n) (1-s) ds.$$
\[ E_n V_0(X_{n+1}, n+1) - V_0(X_n, n) \]
\[ = E_n \sum_{n+1}^\infty a_j V'_x(X_{n+1}) E_{n+1}[h(X_{n+1}, \xi_j) - h(X_{n+1})] \]
\[ - \sum_{n+1}^\infty a_j V'_x(X_n) E_n[h(X_n, \xi_j) - h(X_n)] \]
\[ - a_n V'_x(X_n)[E_n h(X_n, \xi_n) - h(X_n)], \quad (2.3) \]

which equals

last line of (2.3) + \[ a_n E_n h'(X_n, \xi_n) \left\{ \prod_{j=n+1}^\infty a_j E_{n+1} V'_x(X_n + s\delta X_n) \right\} \]
\[ \times (h(X_n + s\delta X_n, \xi_j) - h(X_n + s\delta X_n)) \] \[ \times ds. \quad (2.4) \]

The last term in (2.4) is bounded by \( O(a_n^2) O(1 + |V'_x(X_n) h(X_n)|) \). Define \( \bar{V}(n) = V(X_n) + V_0(X_n, n) \). Then, by the above calculations,

\[ E_n \bar{V}(n+1) - \bar{V}(n) = a_n (1 + a_n \varepsilon_n) V'_x(X_n) h(X_n) + \varepsilon_n a_n^2, \quad (2.5) \]

where \( \{\varepsilon_n\}, \{\varepsilon_n\} \) are sequences of uniformly bounded random variables. Thus we can write

\[ \bar{V}(n) - \sum_{i=0}^{n-1} a_i (1 + a_i \varepsilon_i) V'_x(X_i) h(X_i) - \sum_{i=0}^{n-1} \varepsilon_i a_i^2 = \sum_{i=0}^{n-1} m_i \equiv M_n, \quad (2.6) \]

where (2.6) defines \( m_i, M_n \), and \( \{M_n\} \) is a martingale. Note that

\[ m_n = \bar{V}(n+1) - \bar{V}(n) - a_n (1 + a_n \varepsilon_n) V'_x(X_n) h(X_n) - \varepsilon_n a_n^2. \]

Henceforth, let \( n \) be large enough so that \( |\varepsilon_n a_n| \leq \varepsilon_0/4, \left| a_n \varepsilon_n \right| \leq 1/4 \). Note that \( \bar{V}(n) \geq O(a_n) \) for large \( n \) by (A3), (A5).

Let \( n_0 \) be a stopping time such that \( X_{n_0} \in Q_0 \) and define \( n_1 = \min\{n: n > n_0, X_n \in Q_0\} \). Then \( \{\bar{V}(n \cap n_1), n \geq n_0\} \) is a super martingale bounded below by \(-O(a_n)\), and \( E_n \bar{V}(n+1) - \bar{V}(n) \leq -\varepsilon_0 a_n/2 \) if \( X_n \in Q_0 \) and \( n \) is large. This implies that \( Q_0 \) is a recurrence set; i.e., \( X_n \in Q_0 \) for infinitely many \( n \) w.p.1. Let \( \lambda_1 > \lambda_0 \) and define \( Q_1 = \{x: V(x) \leq \lambda_1\} \). For each such \( Q_1 \) there is a real \( K(Q_1) \) such that \( |m_n|^2 \leq K(Q_1) a_n^2 \) if \( X_n \in Q_1 \). Define \( n_2 = \min\{n: X_n \in Q_1, n \geq n_0\} \). Then

\[ P_{n_0} \sup_{n_0 < n < n_2} \left| \sum_{i=n_0}^{n_1} m_i \right| \geq \varepsilon \leq K(Q_1) E_{n_0} \sum_{i=n_0}^{n_1-1} a_i^2/\varepsilon^2. \quad (2.7) \]

From the above part of this paragraph and the fact that \( V'_x(x) h(x) \leq -\varepsilon_0 \).
for $x \in Q_0$ and the boundedness of $|h(x, \xi)|$, $x \in Q_1$, we conclude that eventually (w.p.1) $X_n$ stays in $Q_1$ (for any $\lambda_1 > \lambda_0$). Also,

$$\sup_{m \geq n} \left| V(X_m) - V(X_n) - \sum_{i=n}^{m-1} a_i(1 + a_i \varepsilon_i) V'_i(X_i) \tilde{h}(X_i) \right| \rightarrow 0$$

w.p.1 as $n \rightarrow \infty$ \hspace{1cm} (2.8a)

or, equivalently, using $m(t_n) = n$,

$$\sup_{s \geq 0} \left| V(X^n(s)) - V(X^n(0)) - \sum_{i=n}^{m(t_n + s) - 1} a_i(1 + a_i \varepsilon_i) V'_i(X_i) \tilde{h}(X_i) \right| \rightarrow 0$$

w.p.1 as $n \rightarrow \infty$. \hspace{1cm} (2.8b)

Let $\Omega = \{\text{set of non-recurrence of } Q_0\} \cup \{\text{set of non-convergence of } \Sigma m_n\}$. By the w.p.1 boundedness of $\{X_n\}$, $X^0(\cdot)$ is uniformly continuous for $\omega \in \Omega$. Fix $\omega \in \Omega_1 \cup \Omega_2 = \Omega_0$. Via the Arzelà–Ascoli Theorem, pick a convergent subsequence (converging uniformly on bounded intervals) of $\{X^n(\cdot)\}$, with limit $X(\cdot)$. Then

$$V(X(t)) = V(X(0)) + \int_0^t V'_i(X(s)) \tilde{h}(X(s)) \, ds.$$ \hspace{1cm} (2.9)

Equation (2.8) implies that if $V'_i(x) \tilde{h}(x) \leq 0$ for all $x$, then $X_n \rightarrow S_0 = \{x : V'_i(x) \tilde{h}(x) = 0\}$ w.p.1 as $n \rightarrow \infty$.

Next, let $f(\cdot)$ be a real valued function on $R'$ with compact support and continuous second derivatives. With $f(\cdot)$ replacing $V(\cdot)$, define $f_0(x, n)$, $\tilde{f}(n)$ as $V_0(x, n)$, $\tilde{V}(n)$ were defined. Then (2.8) holds for $f(x)$ replacing $V(\cdot)$. By choosing $f(\cdot)$ such that $f(x) = x^i$, $i = 1, \ldots, r$, in the set $Q_1$, where $x^i$ is the $i$th component of $x$, we see there is a bounded sequence $\{\varepsilon_n\}$ such that

$$\sup_{s \geq 0} \left| X^n(s) - X^n(0) - \sum_{i=n}^{m(t_n + s) - 1} a_i(1 + a_i \varepsilon_i) \tilde{h}(X_i) \right| \rightarrow 0$$

w.p.1 as $n \rightarrow \infty$. \hspace{1cm} (2.10)

Thus any limit $X(\cdot)$ of $\{X^n(\cdot)\}$ must satisfy (2.1) and the possible limit points of $\{X_n\}$ are contained w.p.1 in the largest bounded invariant set of (2.1). The assertion concerning asymptotically stable $x(t) \equiv x_0$ is now readily proved (see, e.g., proof of Theorem (2.3.1) of [1]), and the details are omitted. \hspace{1cm} Q.E.D.

**Remark.** Note, for future reference in the “unbounded noise” case, that if $\{X_n\}$ were bounded w.p.1 and $a_i \varepsilon_i \rightarrow 0$ w.p.1, then (2.10) implies that any
subsequence of \( \{X^n(\cdot)\} \) has a further subsequence which converges w.p.1 to a continuous function \( X(\cdot) \) satisfying \( \dot{X} = \tilde{h}(X) \). Boundedness of \( h(X,\cdot) \) in \( Q_1 \) is not actually needed.

3. The Projection Method

Let \( G \) be as defined in Section 1. For the continuous vector field \( \tilde{h}(\cdot) \) define \( \tilde{\pi}(\tilde{h}(x)) = \) projection of \( \tilde{h}(x) \) onto \( G \); i.e., \( \tilde{\pi}(\tilde{h}(x)) = \lim_{\Delta \to 0} \left[ \pi_G(x + \Delta \tilde{h}(x)) - x \right] / \Delta \). The limit need not be unique. We will need

(A8) (A3) and (A6) hold, but with \( V_x \) dropped and the right sides \( O(a_n) \).

(A9) \( q_i(\cdot), i = 1,\ldots, m, \) are continuously differentiable, \( G \) is bounded and is the closure of its interior \( G^0 = G - \partial G = \{x: q_i(x) < 0, i = 1,\ldots, m\} \), at each \( x \in \partial G \), the gradients of the active constraints are linearly independent.

**Theorem 2.** Assume (A1), (A8), (A9). Then \( \{X^n(\cdot)\} \) is uniformly continuous on \([0, \infty)\). There is a null set \( \Omega_0 \) such that for \( \omega \in \Omega_0 \) any limit \( X(\cdot) \) of a convergent (uniformly on bounded internals) subsequence of \( \{X^n(\cdot)\} \) satisfies

\[
\dot{x} = \tilde{\pi}(\tilde{h}(x)).
\]

If \( \{X_n\} \subseteq \text{compact } A \subseteq DA(x_0) \) infinitely often and \( \omega \in \Omega_0 \), and \( x_0 \) is an asymptotically stable point of (3.1), then \( X_n \to x_0 \) w.p.1. Let \( H(\cdot) \geq 0 \) be a real valued function whose second mixed partial derivatives are continuous and \( \tilde{h}(x) = -H_x(x) \). Define \( KT = \) set of points where \( \tilde{h}(x) \tilde{\pi}((\tilde{h}(x)) = 0 \), and suppose that \( KT = \bigcup_{i=1}^l S_i \), where the \( S_i \) are disjoint, closed and such that \( H(x) \) is constant on each \( S_i \). Then \( X_n \to KT \) w.p.1 as \( n \to \infty \).

**Proof.** The proof is very similar to that of Theorem 1. Let \( f(\cdot) \) be an arbitrary real valued function on \( R^r \) with continuous second partial derivatives. Then

\[
E_n f(X_{n+1}) - f(X_n) = a_n f'_x(X_n) E_n h(X_n, \xi_n) + a_n f'_x(X_n) E_n \tau_n
\]

\[
+ a_n^2 \int_0^1 E_n(\delta X_n/a_n) f_{xx}(X_n + s\delta X_n)(\delta X_n/a_n)(1 - s) ds,
\]

where \( \tau_n = \pi_G(X_n + a_n h(X_n, \xi_n)) - (X_n + a_n h(X_n, \xi_n)) / a_n = O(1) \). Note that there is a \( K \) such that \( \tau_n = 0 \) if distance \( (X_n, \partial G) \geq Ka_n \) and that \( \tau_n \) lies in the cone \( -C(X_{n+1}) \) where \( C(x) = \{y: y = \sum_i \lambda_i q_{i,x}(x), \lambda_i \geq 0, \text{ and the sum is over the active constraints at } x\} \).
Define \( f_0(x, n) \) by
\[
f_0(x, n) = \sum_{j=n}^{\infty} a_j f'_x(x) E_n [h(x, \xi_j) - \tilde{h}(x)]
\]
and set \( \tilde{f}(n) = f(X_n) + f_0(X_n, n) \). There is a bounded sequence \( \epsilon_i \) such that
\[
E_n \tilde{f}(n + 1) - \tilde{f}(n) - \epsilon_n a_n^2 - a_n f'_x(X_n) \tilde{h}(X_n) - a_n f'_x(X_n) E_n \tau_n = 0.
\]
\[
\tilde{f}(n) - \tilde{f}(0) - \sum_{i=0}^{n-1} \epsilon_i a_i^2 - \sum_{i=0}^{n-1} a_i f'_x(X_i) \tilde{h}(X_i) - \sum_{i=0}^{n-1} a_i f'_x(X_i) \tau_i
\]
\[
= \sum_{i=0}^{n-1} m_i = M_n,
\]
where \( |M_n| \) is a martingale and \( |m_i|^2 \leq Ka_i^2 \). As in Theorem 1,
\[
\sup_{s \geq 0} \left| f(X^n(s)) - f(X^n(0)) - \sum_{i=n}^{m(t_n + s) - 1} a_i f'_x(X_i) \tilde{h}(X_i) - \sum_{i=n}^{m(t_n + s) - 1} a_i f'_x(X_i) \tau_i \right| \to 0
\]
w.p.1 as \( n \to \infty \). (3.2)

from which follows
\[
\sup_{s \geq 0} \left| X^n(s) - X^n(0) - \sum_{i=n}^{m(t_n + s) - 1} a_i \tilde{h}(X_i) - \sum_{i=n}^{m(t_n + s) - 1} a_i \tau_i \right| \to 0
\]
w.p.1 as \( n \to \infty \). (3.3)

Also, \( \{X^n(\cdot)\} \) is equicontinuous, since \( h(\cdot, \cdot) \) is bounded.

Let \( \Omega_0 \) denote the set of nonconvergence in (3.3) and for fixed \( \omega \in \Omega_0 \),
extract a convergent subsequence of \( \{X^n(\cdot)\} \) (uniformly on bounded intervals) with limit denoted by \( X(\cdot) \). Define \( \tilde{h}_0(x) = \tilde{\pi}(\tilde{h}(x)) \) and \( \tilde{h}_1(x) = \tilde{h}(x) - \tilde{h}_0(x) \). Then, by (3.3) there is a bounded \( R^\tau \)-valued measurable function \( \tau(\cdot) \) such that \( \tau(s) = 0 \) unless \( X(s) \in \partial G \), and if \( X(s) \in \partial G \) then \( \tau(s) \) is in the cone \( -C(X(s)) \) and (3.4) holds:

\[
X(t) = X(0) + \int_0^t \tilde{h}_0(X(s)) \, ds + \int_0^t \tau(s) \, ds
\]

(3.4)

The last two integrals on the right of (3.4) must cancel if \( X(t) \) is to remain in \( G \) for all \( t \). Thus (3.1) holds w.p.1.

If \( \tilde{h}(x) = -H_x(x) \), then use \( \Pi(\cdot) \) as a Liapunov function for (3.1) to get
\[
\tilde{H}(x) = H'_x(x) \tilde{\pi}(-H_x(x)) \leq 0,
\]
(3.5)
from which we see that \( X(t) \to KT \) as \( t \to \infty \). Thus, for each \( \varepsilon > 0 \), \( \{ X_n \} \) is in an \( \varepsilon \) neighborhood \( N_\varepsilon(KT) \) of \( KT \) infinitely often w.p.1. Fix \( \varepsilon > 0 \). Define 
\[ H_1 = \lim_{n \to \infty} H(X_n). \]
Suppose that \( S_1 \) and \( \hat{\Omega}_1 \) are such that \( H_1 = \) value of \( H(x) \) on \( S_1 \) if \( \omega \in \hat{\Omega}_1 \) and \( P(\hat{\Omega}_1) > 0 \), and for some \( \varepsilon_i > \varepsilon > 0 \), \( \{ X_n \} \) leaves the \( \varepsilon_i \)-neighborhood \( N_\varepsilon(S_1) \) infinitely often for \( \omega \in \hat{\Omega}_1 \). Then for (almost all) \( \omega \in \hat{\Omega}_1 \), there are real numbers \( I_n \to \infty \) and \( k_n \geq K_0 > 0 \) with \( k_n \to T \leq \infty \) and a solution \( X(\cdot) \) to (3.1) which is a limit of the sequence \( \{ X^n(l_n + s), s < k_n \} \) for \( m = 1, 2, \ldots \).

Using an argument like that used in [1], Theorem 2.3.5, the last sentence and (3.5) imply that 
\[ H_1 \neq \lim_{n \to \infty} H(X_n) \] almost everywhere on \( \hat{\Omega}_1 \), a contradiction. The next to the last assertion of the theorem is proved in a similar way. Q.E.D.

4. STATE DEPENDENT AND UNBOUNDED NOISE;
STATE DEPENDENT AND BOUNDED NOISE

There are several ways in which the state dependent and bounded noise case can be treated. The noise can be parameterized as in [4, Section 9]. Here, we choose a Markovian representation. Suppose that \( \{ \xi_{n-1}, X_n \} \) is a Markov process. In applications, this might require an augmentation of the state space of the “original” \( \{ \xi_n \} \) and a redefinition of the “original” \( h(\cdot, \cdot) \).

Let \( E_n \) denote conditioning on \( \xi_j, j < n, X_j, j \leq n \), and define the “partial” transition function \( P(\xi, \alpha, \Gamma|x) = P(\xi_n \in \Gamma|X_m = x, \xi_{n-1} = \xi) \) and for \( m > 1 \), define \( P(\xi, m, \Gamma|x) \) by

\[ P(\xi, \alpha + \beta, \Gamma|x) = \int P(\xi, \alpha, dy|x) P(y, \beta, \Gamma|x). \]

It is supposed that \( P \) does not depend on \( n \), for notational simplicity only.

Write \( V_\alpha(x, n) \) in the form

\[ V_\alpha(x, n) = V'_x(x) \sum_{j=n}^{\infty} a_j \int h(x, \xi) P(\xi_{n-1}, j - n + 1, d\xi|x) - \bar{h}(x). \]  

Note that \( E_n P(\xi_n, j - n, \Gamma|X_n) = P(\xi_{n-1}, j - n + 1, \Gamma|X_n) \) by the Markov property. Assume that the sum in (4.1) is continuously differentiable in \( x \), and that the derivatives can be taken termwise and that (replacing A6))

\[ \left| \sum_{j=n+1}^{\infty} a_j \left[ V'_x(x) \int h(x, \xi) P(\xi_n, j - n, d\xi|x) - \bar{h}(x) \right] \right| \leq Ka_n(1 + |V'_x(x) \bar{h}(x)|^{1/2}). \]  

(4.2)
THEOREM 3. Assume (A1)–(A7) but with (4.1), (4.2) replacing (A3), (A6), resp. Then the conclusions of Theorem 1 hold.

Assume (A1), (A8), (A9) but with the modifications of (A3), (A6) stated above. Then the conclusions of Theorem 2 continue to hold.

Remark on the proof. In the proof the difference (4.3) occurs,

\[
\sum_{j=n+1}^{\infty} E_a a_j V'_x(X_{n+j}) \left[ h(X_{n+j}, \xi) P(\xi_n, j - n, d\xi | X_{n+j}) - \bar{h}(X_{n+j}) \right] 
\]

Using the differentiability and the equality below (4.1) and the bounds from (A1)–(A7) (modified for Theorem 3), (4.3) can be seen to be of the order of \( 41 + I K_k W(a) \).

The proof of Theorem 3 is the same as those of Theorems 1 and 2.

Unbounded noise. We state a generalization of Theorem 1 for the case where \( \{\xi_n\} \) is unbounded. First, make the following alterations in the assumptions. Drop the boundedness of \( \{\xi_n\} \) in (A1). Suppose that there are random sequences \( \{\alpha_{in}\} \) with the following properties. For each \( Q_i \supset Q_0 \), there is \( \{\alpha_{on}\} \) such that \( \|h(x, \xi_n)\| \leq \alpha_{on}, x \in Q_i \) and \( \sum E\alpha_{on}^2 \leq \infty \). Also \( E_n \|h(x, \xi_n)\|^2 \leq \alpha_{on}^2 (1 + |V'_x(x) h(x)|) \), \( \alpha_{on}^2 a_{in} \to 0 \) w.p.1, and \( |V_0(x, n)| \leq \alpha_{on} (1 + |V'_x(x) \bar{h}(x)|) \), \( \sum E\alpha_{on}^2 < \infty \). Both (A6) and (A7) were used to get the bound below (2.4) on (2.4). We require that the bound hold with \( O(\alpha_{on}^2) \) replaced by \( \alpha_{on}^2 a_{in} \), where \( \sup_a E\alpha_{on}^2 \alpha_{on}^3 < \infty \). This is, perhaps, an awkward way of stating the assumption, but it can be verified in many standard examples. We require also that these alterations hold when \( f(\cdot) \in \mathcal{E}_0^a \) replaces \( V(\cdot) \), although the \( \alpha_{on} \) might depend on \( f(\cdot) \). We now have

THEOREM 4. Under the conditions of Theorem 1, altered as above, the conclusions of Theorem 1 continue to hold.

The proof is very similar to that of Theorem 1, and only a few changes are required. The \( a_n e_n, a_n \bar{e}_n, a_n \delta_n \) are replaced by sequences which tend to zero w.p.1 as \( n \to \infty \). Also, \( P(n) \geq -K a_{in} a_n \uparrow 0 \) w.p.1 and \( m_n^2 \leq K(Q_1) a_n^2 \sum_{l=0}^{3} \alpha_{in}^2 \) for some real number \( K(Q_1) \).

Remark. All the foregoing results hold if \( \{a_{in}\} \) is random, under the
following additional conditions. \( a_n \) depends on \( \{X_i, i \leq n\} \) only, \( \sum_n a_n = \infty \), \( \sum_n a_n^2 < \infty \), \( \sum_n |a_{n+1} - a_n| < \infty \) w.p.1 and with

\[
\hat{V}_0(x, n) = a_n \sum_{j=n}^{\infty} E_n V^r(x) |h(x, \xi_j) - \bar{h}(x)|
\]

(4.4)

replacing the \( V_0(x, n) \) of (A3).

The "projected" form of Theorem 4 is:

**Theorem 5.** Assume (A9), (A1). Assume (A3), (A6), resp., but with arbitrary \( f(\cdot) \in \mathcal{F}_0^2 \) only and, resp., \( \alpha_1 K \) and \( \alpha_2 K \) replacing \( K \). Assume \( |h(x, \xi_n)| \leq Ka_{on}, x \in G \) and the bound below (2.4) with \( V(\cdot) \) replaced by \( f(\cdot) \in \mathcal{F}_0^2, \) and \( K \) by \( \alpha_3 K \). Assume the conditions on \( \{a_{in}\} \) above Theorem 4. Then any subsequence of \( \{X_n(\cdot)\} \) has a further subsequence which converges uniformly w.p.1 on bounded intervals to a solution of (3.1).

Under the additional assumptions below (3.1), the conclusions there continue to hold.

**Example.** Let \( \{\xi_n\} \) be stationary and Markov and \( h(x, \xi) = \bar{h}(x) + h_0(x) g(\xi) \), where \( E_g(\xi) = 0, E_g^2(\xi) < \infty \). Such a form occurs in applications to the identification and adaptive control of linear systems, where \( \bar{h} \) and \( h_0 \) are affine functions of \( x \). Then, Theorem 1 holds under a simple stability condition on \( \bar{x} = \bar{h}(x) \), and on reasonable conditions on \( \{\xi_n\} \). A standard and important special case occurs in the identification problem for linear systems where we use \( \psi_n = L_1 \xi_n, y_n = L_2 \xi_n, \{\xi_n\} \) Markov and

\[
X_{n+1} = X_n - a_n \psi_n [\psi_n' X_n - y_n],
\]

\( \psi_n \subset R^n, y_n \subset R \).

The following two classes of examples have state dependent noise and they illustrate two different ways of using Theorem 3.

5. A LEARNING AUTOMATA EXAMPLE

This example is a modification of one in [5], where \( a_n \equiv \varepsilon > 0 \) and an extensive development of the asymptotic distributional properties is given. Here we are concerned with w.p.1 convergence only for the case where \( a_n \to 0 \). A relatively simple case is treated. Clearly, more complicated arrival and adaptive processes and systems can be treated.

The problem. Calls arrive at a switching terminal at random at time instants \( n = 0, 1, 2, \ldots \), with \( P\{ \text{one call arrives at } n\text{th instant} \} = \mu \in (0, 1) \),
$P\{\text{> 1 call arrives at } n\text{th instant}\} = 0$. There are two possible routings to the
destination, routes $i$, $i = 1, 2$, where route $i$ has $N_i$ independent lines—and
can handle up to $N_i$ calls simultaneously. Let $(n, n + 1)$ denote the $n$th
interval of time. The duration of each call has the distribution: $P\{\text{call}
completed in the } (n + 1)\text{st interval} |\text{uncompleted at end of } n\text{th interval, route}
i \text{used}\} = \lambda_i \in (0, 1)$. The members of the sequence of interarrival times and
call durations are mutually independent. The use of an adaptive automaton
for adjusting the routing comes from [7].

The routing automaton operates as follows. Let $\{X_n\}$ denote a sequence of
random variables—with values in $[0, 1]$. In order to have an unambiguous
sequencing of events, let the calls ending in the $n$th interval actually end at
time $n + 1/2$, and let both arrivals and route assignments be at the ends of
the intervals; i.e., at the instants $0, 1, 2, \ldots$, precisely. Thus the state of the
route occupancy at time $(n + 1)^-$ does not include the calls just terminated
or calls arriving at $(n + 1)$. Define the “route occupancy process” $Z_n =
(Z_n^1, Z_n^2)$, where $Z_n^i$ is the number of lines of route $i$ occupied at time $n^+$. 
Thus, $Z_n^i \leq N_i$. If a call arrives at instant $n + 1$, the automaton chooses route
1 with probability $X_n$ and route 2 with probability $1 - X_n$. 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. The choice probabilities $\{X_n\}$ are to be adjusted or
adapted according to the “experience” of the system.

The specific adjustment scheme for $\{X_n\}$ is the following “linear-reward”
algorithm [7]. 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 $i$\}. For practical as well
as theoretical purposes, it is important to bound $X_n$ away from the points 0
and 1. Let $0 < x_1 < x_u < 1$. We use the (projected) algorithm (5.1), where $|x_i^s$
denotes truncation at $x_u$ or $x_1$, and $\alpha(x) = 1 - x, \beta(x) = x$.

$$X_{n+1} = [X_n + a_n \alpha(X_n) J_{1n} + a_n \beta(X_n) J_{2n}]|_{x_i^s}.$$ (5.1)

Some definitions. If the choice probabilities $X_n$ are held fixed at some
value $x$ for all $n$, then the route choice automaton still is well defined. For
fixed route selection probability $x \in (0, 1)$, let $\{Z_n(x)\} = \{(Z_n^1(x), Z_n^2(x)),$
$0 \leq n < \infty \}$ denote the corresponding route occupancy process. For the
process $\{Z_n(x)\}$, the state space $Z = \{(i, j); i \leq N_1, j \leq N_2\}$ (whose points are
ordered in some fixed way) is a single ergodic class, and the probability tran-
sition matrix, denoted by $A'(x)$, has infinitely differentiable components.
With given initial condition, define $P_n(\alpha| x) = P\{Z_n(x) = \alpha\}$ and define the
vector $P_n(x) = \{P_n(\alpha| x), \alpha \in Z\}$. Then $P_{n+1}(x) = A(x) P_n(x)$.

The pair $\{(Z_n, X_n), n \geq 0\}$ is a Markov process on $Z \times [x_1, x_u]$ and the
marginal transition probability $P\{Z_{n+1} = (k, l) | Z_n = (i, j), X_n\}$ is just the
($(i, j)$-column, $(k, l)$-row) entry of $A(X_n)$. Define the vector
\( P_n = \{ P_n(\alpha), \alpha \in Z \} \) where \( P_n(\alpha) = P\{ Z_n = \alpha \mid X_1, \ldots, X_l \} \). Then \( P_{n+1} = A(X_n) \, P_n \). Also, let \( P(x) = \{ P(\alpha \mid x), \alpha \in Z \} \) denote the unique invariant measure for \( \{ Z_n(x) \} \), with marginal defined by \( P^1(N_1 \mid x) = \) asymptotic probability that \( Z_n = N_1 \), and similarly for route 2. Finally, define the transition probability \( P(\alpha, j, \alpha_1 \mid x) = P\{ Z_j(x) = \alpha, Z_{j+1}(x) = \alpha_1 \mid x \} \), and define the marginal transition probability

\[
P^1(\alpha, j, N_i \mid x) = P\{ Z_j(x) = N_i \mid x \} = P^1(\alpha_1 \mid x).
\]

Define \( E_n \) to be the expectation conditioned on \( \{ Z_l, X_l \mid l \leq n \} \) and set \( v_i = (1 - \lambda_i)^X_i \).

**Application of Theorem 3.** We have \( h(X_n, \xi_n) = \alpha(X_n) J_1 + \beta(X_n) J_2 \) and, with \( I\{ \cdot \} \) denoting the indicator function,

\[
E_n h(X_n, \xi_n) = \mu \alpha(X_n) X_n \{ 1 - v_1 I\{ Z_n = N_1 \} \} + \mu \beta(X_n) (1 - X_n) \{ 1 - v_2 I\{ Z_n = N_2 \} \}.
\]

which can be written in the form

\[
-\mu X_n (1 - X_n) [v_2 P^1(Z_n, 0, N_2 \mid X_n) - v_1 P^1(Z_n, 0, N_1 \mid X_n)]. \tag{5.2}
\]

Define \( \overline{h}(\cdot) \) to be the limit

\[
\overline{h}(x) = \mu x (1 - x) \lim_{n \to \infty} E[v_2 P^2(Z, n, N_2 \mid x) - v_1 P^1(Z, n, N_2 \mid x)]. \tag{5.3}
\]

The sum (A3) is replaced by (since the second part of Theorem 3 is to be used, the \( V_x(x) \) component can be dropped)

\[
V_0(x, n) = \mu x (1 - x) \overline{V}_x(x) \sum_{j=n}^{\infty} a_j [v_2 (P^2(x, j - n, N_2 \mid x) - P^2(N_2 \mid x)) - v_1 (P^1(x, j - n, N_1 \mid x) - P^1(N_1 \mid x))]. \tag{5.4}
\]

The sum (A6) is replaced by the analogous sum of the derivatives (again drop the \( V_x(x) \) component). There is a unique \( \bar{x} \in (0, 1) \) such that \( \overline{h}(\bar{x}) = 0 \) and \( \overline{h}(x) > 0 \) for \( x \in (0, \bar{x}) \) and \( \overline{h}(x) < 0 \) for \( x \in (\bar{x}, 1) \). The \( P_n(x) \) and \( P_{n,x}(x) \) converge [5] to the limits \( P(x) \), \( P_x(x) \) geometrically with a rate uniform in \( x \in [x_1, x_u] \) and in \( P_0(x) \) \( P_{0,x}(x) = 0 \) is the appropriate initial condition to get the limit for the derivatives for use in (A6)). This result implies that (A3), (A6) exist and converge absolutely and uniformly in \( (n, X_n) \) at a geometric rate. See [5] for the details of the convergences.

Part 2 of Theorem 3 now yields Theorem 4 below. Theorem 4 can also be proved directly, via the method of Theorem 2 (here the boundary is only
STOCHASTIC APPROXIMATION

|\{x_l, x_u\}| with the "corrected" test function (5.4) used in lieu of the sum in (A3).

**Theorem 6.** Let \( \Sigma a_i^2 < \infty, \Sigma a_i = \infty \). Then if \( \bar{x} \in [x_l, x_u] \), we have \( |X_n| \to \bar{x} \) w.p.1. Otherwise \( \{X_n\} \) converges w.p.1 to the point \( x_l \) or \( x_u \) which is nearest to \( \bar{x} \).

6. OBSERVATION AVERAGING FOR STOCHASTIC APPROXIMATIONS

The general method of Theorems 1 and 2 can be easily used to prove w.p.1 convergence for stochastic approximations of the Robbins–Monro or Kiefer–Wolfowitz type but with averaged observations. The main difficulty is due to the fact that the quantity which plays the role of the noise is always state dependent. The idea will be illustrated via a very simple example. We use a Robbins–Monro scheme to estimate the root of \( Kx = 0, x = \text{scalar}, K > 0 \) (but the method is applicable to the general problem).

Define the estimates by

\[
X_{n+1} = (X_n + a_n \xi_n/x_n)_{x_n},
\]

\[
\xi_n = a \xi_{n-1} - \beta[KX_n + \psi_n],
\]

where \( a \in (0, 1) \), \( \beta > 0 \), and \( \{\psi_n\} \) is a bounded sequence of mutually independent random variables with zero mean value. If \( a = 0 \), then (6.1) is the usual Robbins–Monro method, truncated at values \( x_l, x_u \). If \( a \in (0, 1) \), then the observations are exponentially weighted. Theorem 3 requires truncation to some finite interval \([x_l, x_u]\). Such truncation is usually done in practice anyway. Define \( h(x) = -\beta Kx/(1 - a) \) and \( h(x, \xi) = \xi \). Instead of writing \( V_0(x, n) \) in the form (4.1), it is more convenient to do the following. For each \( x, n \), define the auxiliary processes \( \{\xi_j(x), j \geq n\} \) where the initial condition \( \xi_{n-1}(x) \) is to be defined and \( \xi_j(x) = a \xi_{j-1}(x) - (\beta Kx + \psi_j), j \geq n \).

Write \( V_0(x, n) \) as

\[
V_0(x, n) = \sum_{j=n}^\infty a_j V'_{\xi_j}(x) E_n[h(x, \xi_j(x)) - h(x)],
\]

where \( \xi_{n-1}(x) = \xi_{n-1} \), and \( E_n \) denotes expectation conditioned on \( X_i, i < n, \psi_j, i < n \). Note that \( \xi_n(X_n) = \xi_n \).

Now Theorem 3 yields

**Theorem 7.** Let \( \Sigma a_i^2 < \infty, \Sigma a_i = \infty \). If \( 0 \in [x_l, x_u] \), then \( \{X_n\} \to 0 \) w.p.1. Otherwise \( \{X_n\} \) converges w.p.1 to the point \( x_l, x_u \) which is closest to zero.

In [4] there is an analysis of the asymptotic properties of (6.1) when \( a_n \equiv \varepsilon > 0 \).
7. Weak Convergence

Theorem 8. Assume the conditions of Theorem 4, except let $a_n \to 0$ replace $\Sigma a^2_n < \infty$. Let $S$ denote the largest bounded invariant set of (2.1). Then any subsequence of $\{X^n(\cdot)\}$ contains a further subsequence which converges w.p.1 to a solution of (2.1) in $S$. Also $X_n \to S$ in probability. More strongly, for each $T < \infty$, $\varepsilon > 0$ ($d(\cdot, \cdot)$ denotes distance)

$$
\lim_{n \to \infty} P \left\{ \sup_{t \leq T} d(X^n(t), S) \geq \varepsilon \right\} = 0. \tag{7.1}
$$

If (3.1) and the conditions of Theorem 5 replace (2.1) and the conditions of Theorem 4 (with $a_n \to 0$ replacing $\xi a^2_n < \infty$), then the conclusions remain valid for the projection algorithm.

Proof: Part 1. Recurrence of $Q_0$. Suppose that the $a_n a_{nl} \to 0$ uniformly in $\omega$ as $n \to \infty$. Then the proof in Theorem 1 that $Q_0$ is a recurrence set (above (2.7), and suitably modified for the "unbounded" noise conditions of Theorem 4) implies that $Q_0$ is a recurrence set here. But, for any $\varepsilon > 0$, there is an $n_\varepsilon < \infty$ such that $P\{a_n \sum_{i=0}^{\infty} a_{nl} > \varepsilon$, some $n \geq n_\varepsilon \} \leq \varepsilon$. Thus, for large enough $n$, we can suppose that the $a_n a_{nl}$ are as small as desired, uniformly in $\omega$, by modifying the process on a set of arbitrarily small probability. The foregoing statements imply that $Q_0$ is a recurrence set. For the purposes of the proof, we can (and will) continue to suppose that the $a_n a_{nl} \to 0$ uniformly in $\omega$.

Part 2. Tightness of $\{X_n\}$. We have $a_n \sup_{x \in Q_1} |h(x, \xi_n)| \to 0$ w.p.1 as $n \to \infty$, and we suppose that this quantity is as small as desired for large enough $n$. Then, if $X_n \in Q_0$, we can suppose that $X_{n+1} \in Q_1$. Let $\tau \leq \infty$ denote any random time such that $X^0(\tau) \in Q_1 - Q_0$; in particular, let $\tau = \min\{t: X^0(t) \in Q_1 - Q_0, t \geq t_m\}$ for some large $m$. Let $t_\tau = \min\{t: X^0(t) \in Q_0, t > \tau\}$. Then, for large enough $n$, $E[t_{\tau} - \tau | X^0(s), s \leq \tau] \leq 2\lambda_1/e_0 \equiv T_0$ on the set where $\tau < \infty$. Also, if $\tau$ is large enough (using the supermartingale property of $\{\bar{P}(n)\}$ in $R' - Q_0$)

$$
P \left\{ \sup_{\tau_1 > \tau_0 > \tau} \bar{P}(n) \geq \lambda \right\} \leq \frac{E \bar{P}(m(\tau))}{\lambda} \leq \frac{\lambda_1}{\lambda}. \tag{7.2b}
$$

Now, define the random times $\{\sigma_n\}$ in the following way.

$$
\sigma_1 = \min\{t: X^0(t) \in Q_0\}, \quad \text{and, for } n > 1 \quad \sigma_n = \sigma_{n-1} + T_0 \quad \text{if } X^0(t) \in Q_0 \quad \text{for } t \in [\sigma_{n-1}, \sigma_{n-1} + T_0], \tag{7.2a}
$$

$$
= \inf\{t: t > \sigma_{n-1}, X^0(t) \notin Q_0\} \quad \text{if } X^0(\sigma_{n-1}) \in Q_0,
$$

but (7.2a) does not hold,

$$
\text{but (7.2a) does not hold,} \tag{7.2b}
$$

$$
= \min\{t: t > \sigma_{n-1}, X^0(t) \in Q_0\} \quad \text{if } X^0(\sigma_{n-1}) \in Q_1 - Q_0. \tag{7.2c}
$$
Let $\bar{E}_n$ and $\bar{P}_n$ denote expectation and probability conditioned on $|X^0(s), s \leq \sigma_n|$. Then $\bar{E}_n[\sigma_{n+1} - \sigma_n] \leq T_0$ for large $n$.

Fix $\delta > 0$. For each $t > 0$, and integer $k$, define $j(t, k)$ by

$$j(t, k) = \min\{j: \bar{P}_j[t > \sigma_{k+j} \leq \delta]\}.$$  

Then, for $\lambda > \lambda_1$, and the intervals $A_i$ defined by $A_1 = [\sigma_{j(t, k)}, \sigma_{j(t, k)+1}], A_2 = [\sigma_{j(t, k)+1}, \sigma_{j(t, k)+2}], \ldots$,

$$P[\bar{V}(X^0(t)) \geq \lambda] \leq P[\sigma_{j(t, k)} \geq t] + \sum_{j=1}^{k} P[\sup_{s \in A_i} \bar{V}(X^0(s)) \geq \lambda].$$  \hspace{1cm} (7.3)

Now, choose $k$ such that the first term on the right of (7.3) is $\leq \delta/4$. The choose $\lambda$ such that the second is $\leq \delta/2$. For each $\delta$, there is a $t_\delta < \infty$ such that these choices can be made for all $t > t_\delta$, by the comments in the first paragraph of this part. By (A3) and since $V(x) \to \infty$ as $|x| \to \infty$, we have tightness of $\{X_n\}$.

**Part 3. The weak convergence (7.1).** Suppose that (7.1) is false. Then there is some subsequence $\{X^n(\cdot)\}$ such that no further subsequence satisfies (7.1). We obtain a contradiction to this by showing that if $k_n \to \infty$ fast enough as $n \to \infty$ and $\{X^n(0)\}$ converges weakly, then $\{X^n(\cdot)\}$ converges weakly to a solution $X(\cdot)$ of (2.1), where $X(t) \in S$ for all $t \in (-\infty, \infty)$. Choose $k_n \to \infty$ and $T_n \to \infty$ as $n \to \infty$ such that $t_{k_n} - T_n \to \infty$ and that

$$\sum_{i=m(t_{k_n}+T_n)}^{m(t_{k_n}+T_n)} a_i^2 < \infty.$$  \hspace{1cm} (7.4)

Using (7.4) in an argument like that used in Theorem 1 (suitably modified for the "unbounded noise" case) and the tightness of $\{X_n\}$, we can show that

$$\lim_{n \to \infty} P\{X^n(t) \in Q_0, \text{ some } t \in [-T_n, -T_n/2]\} = 1,$$

$$\lim_{n \to \infty} P\{X^n(t) \in Q_1, \text{ all } t \in [-T_n/2, T_n]\} = 1,$$

$$\lim_{n \to \infty} \sup_{T_n > t > -T_n/2} \left| X^n(t) - X^n(0) - \int_0^t h(X^n(s)) \, ds \right| = 0 \text{ w.p.1.}$$  \hspace{1cm} (7.6)

By Part 2, $\{X^n(0)\}$ is tight. If $\{X^n(0)\}$ converges weakly, then (7.5)–(7.6) imply weak convergence in $D'(\infty, \infty)$ of $\{X^n(\cdot)\}$ to a solution $X(\cdot)$ of (2.1). But $X(\cdot)$ is bounded, since its paths must lie in $Q_0$. This implies that $X(t) \in S$, all $t \in (-\infty, \infty)$. We have obtained the desired contradiction, and the first part of the theorem is proved.

The result for the projection algorithm is proved in a similar way (except, of course, $\{X_n\}$ is already bounded) and we omit the details. Q.E.D.


