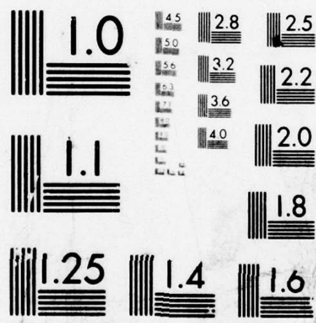
AD-A074 318 BROWN UNIV PROVIDENCE R I LEFSCHETZ CENTER FOR DYNAM--ETC F/G 12/1
ASYMPTOTIC PROPERTIES OF STOCHASTIC APPROXIMATIONS WITH CONSTAN--ETC(U)
APR 79 H J KUSHNER, H HUANG N00014-76-C-0279
UNCLASSIFIED LCDS-TR-79-2 AFOSR-TR-79-0977 NL

1 OF 1
AD
A074318

END
DATE
FILMED
10-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LCDS TR 79-2

AFOSR-TR- 79-0977

HAROLD J. KUSHNER
HAI HUANG

ASYMPTOTIC PROPERTIES OF STOCHASTIC APPROXIMATIONS WITH CONSTANT COEFFICIENTS

APRIL 1979

LEVEL II

Lefschetz Center for Dynamical Systems

Division of Applied Mathematics Brown University Providence RI 02912

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19 REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 18 AFOSR-TR-79-0977
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): 6 ASYMPTOTIC PROPERTIES OF STOCHASTIC APPROXIMATIONS WITH CONSTANT COEFFICIENTS.
5. TYPE OF REPORT & PERIOD COVERED: 9 Interim rept.
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): 10 Harold J. Kushner, Hai Huang
8. CONTRACT OR GRANT NUMBER(s): 15 N00014-76-C-0279, AFOSR-76-3063
9. PERFORMING ORGANIZATION NAME AND ADDRESS:
   Brown University
   Division of Applied Mathematics
   Providence, RI 02912
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 16 61102F 2304/A1
11. CONTROLLING OFFICE NAME AND ADDRESS:
    Air Force Office of Scientific Research/NM
    Bolling AFB, Washington, DC 20332
12. REPORT DATE: 11 Apr[redacted]979
13. NUMBER OF PAGES: 42
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 17 A1
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report):
    Approved for public release; distribution unlimited.
    12 46p.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
    14 LCDS-TR-79-2
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):

In this paper, we obtain analogous results concerning asymptotic behavior and rate of convergence for the case where $a_n = a$, a small constant. The algorithm will be written in the form (1.2), where f and g are measurable funtions.

401 834 Gue

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ASYMPTOTIC PROPERTIES OF STOCHASTIC APPROXIMATIONS
WITH CONSTANT COEFFICIENTS

Harold J. Kushner[+] and Hai Huang[++]

Lefschetz Center for Dynamical Systems
Division of Applied Mathematics
Brown University
Providence, Rhode Island 02912

Accession For	
NTIS GRA&I	☑
DDC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist. A	Avail and/or special

April, 1979

[+]This research was supported, in part by the Air Force Office of Scientific Research under AF-AFOSR 76-3063, in part by the National Science Foundation under NSF-Eng 77-12946, and in part by the Office of Naval Research under N0014-76-C-0279-P0002. OK

[++]This research was supported in part by the Air Force Office of Scientific Research under AF-AFOSR 76-3063.

ASYMPTOTIC PROPERTIES OF STOCHASTIC APPROXIMATIONS WITH CONSTANT COEFFICIENTS

Harold J. Kushner and Hai Huang

Abstract: Asymptotic properties (as $a \to 0$, $n \to \infty$) of the Stochastic Approximation (SA) algorithm

$$(*) \qquad X_{n+1} = X_n + ah(X_n, \xi_n)$$

are obtained, where h is not necessarily additive in ξ_n. If $Eh(x,\xi_n) = g(x)$ and $\dot{x} = g(x)$ is asymptotically stable about a solution $x_t = \theta$, then the asymptotic properties of $\{(X_n - \theta)/\sqrt{a}\} \equiv \{U_n^a\}$ are developed. In particular, it is shown that (as $a \to 0$) a natural continuous parameter interpolation of $\{U_n^a\}$ converges weakly to a linear diffusion process, from which the asymptotic properties of $\{U_n^a\}$ and $\{X_n\}$ for small a can be obtained. The conditions on $\{\xi_n\}$ are reasonable from the point of view of the usual applications to adaptive systems and identification. These results seem to be the first of their type for SA's with constant coefficients. Some rate of convergence results for classical SA's are improved. Also, an application of (*) to a problem of tracking the time varying parameters of a linear system is discussed, and a limit theorem obtained. Because in the usual practical implementations of SA to problems in systems theory, the gain sequence $\{a_n\}$ does not normally go to zero (due to considerations of robustness and non-stationarities), these results are of particular importance.

ASYMPTOTIC PROPERTIES OF STOCHASTIC APPROXIMATIONS
WITH CONSTANT COEFFICIENTS

1. Introduction

In [1] rates of convergence for stochastic approximations (SA) of the type

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

were treated, where $\{a_n\}$ is a sequence of positive numbers tending to zero and such that $\Sigma\, a_n = \infty$, and $\{\xi_n\}$ is a sequence of random variables. In particular, we used $a_n = A/(n+1)^{\alpha}$, $0 < \alpha \leq 1$, although the proofs could have been adapted to deal with more general sequences. As has been usual in rate of convergence studies for SA's, it was assumed that there is a vector θ such that $X_n \to \theta$ w.p.1, and that $\{\xi_n\}$ is a stationary sequence. Unlike previous works on the rate of convergence problem, [1] did not assume that h is additive in ξ_n; the additivity assumption is not satisfied by many important applications in systems theory.

In this paper, we obtain analogous results concerning asymptotic behavior and rate of convergence for the case where $a_n = a$, a small constant. The algorithm will be written in the form (1.2), where f and g are measurable functions, further properties of which will be given below.

$$X^a_{n+1} = X^a_n + ah(X^a_n,\xi^a_n) \equiv X^a_n + ag(X^a_n) + af(X^a_n,\xi^a_n)$$

(1.2)

$$X^a_0 = X_0, \text{ independent of } a$$

Algorithms of the type (1.2) are particularly important in applications to both identification theory and adaptive systems theory, and for a version of this problem, the results are both specialized and extended in Sections 6 and 7; in Section 7 $\{\xi^a_n\}$ is non-stationary, and the "parameter tracking problem" is dealt with. In such applications, h is not additive in the noise ξ^a_n, and the $\{\xi^a_n\}$ may not be a stationary sequence. Furthermore, in engineering practice there is usually a constant $a > 0$ such that either $\{a_n\}$ tends to a or else that $a_n \equiv a$, although almost all the existing analysis of (1.1), (1.2) (indeed of all SA methods [2], [3], [4]) assume $a_n \to 0$. The case (1.2) is more robust than (1.1) in the sense that it can better accommodate non-stationarities and modelling errors, and it is often the form used in applications.

In general, little is known about the sequence (1.2). Normally, $\{X^a_n\}$ does not converge w.p.1, and if $\{\xi^a_n\}$ is non-stationary it may not even converge in distribution. Under various assumptions, (1.3) (a specialization of (1.2)) has been treated in the adaptive process literature. Here B is a vector valued bilinear form and A,C are matrices (Widrow et al [5], Senne [6], Davisson [7]).

$$X^a_{n+1} = X^a_n + aB(X^a_n,\xi^a_n) + aC\xi^a_n + aAX^a_n. \tag{1.3}$$

Results such as $\overline{\lim_{n}} E|X_n^a|^2 \to 0$ as $a \to 0$ were obtained. Our method works under broader conditions (we assume bounded noise; for a form of (1.3), [7] dealt with unbounded but m-dependent Gaussian noise and used a trick similar to the concatenation of m steps into one and then exploitation of a result for independent $\{\xi_n^a\}$) and yields a much more complete picture of the process behavior. As in [1], [3], weak convergence methods are used.

Define $U_n^a = (X_n^a - \theta)/\sqrt{a}$ and $t_n = an$. Let N_a be a sequence which goes to ∞ as $a \to 0$, and define the piecewise constant continuous parameter process $U^a(\cdot)$ by $U^a(0) = U_{N_a}^a$, $U^a(t) = U_{n+N_a}^a$ in $[na,(n+1)a)$. We prove that $\{U^a(\cdot)\}$ converges weakly to the Gaussian diffusion (5.1) as $a \to 0$, where R is defined below (5.1) and $\bar{H} = g_x(\theta)$. The results yield stability of the process (1.2) for small a, together with the asymptotic (as $a \to 0$) error variances and correlation functions (of $U^a(\cdot)$). It seems to us that the general approach is quite straightforward and relatively easy to use. The weak convergence and stability ideas yield a lot of intuitive insight into the relations between the structure of an algorithm and its asymptotic properties. Since it makes no sense to assume convergence $X_n \to$ some θ a priori, some stability analysis is needed. For the special adaptive process case when (1.2) reduces to (1.3), the situation is simpler, and we obtain better results in Section 7.

In Section 2, assumptions for the general problem are stated. Tightness of $\{U_n^a, n \geq N_a\}$ is obtained in Section 3.

Section 4 contains some remarks concerning special cases, and on the use of the methods of this paper to extend known convergence and rate of convergence results for SA's of the type (1.1) when the SA sequence converges in probability rather than (the usual assumption) almost surely.

The main limit theorem is given in Section 5, and Section 7 treats the special case (1.3) when the statistics of $\{\xi_n^a\}$ are time varying and both $\bar{H}$ and R are functions of time. Some of the arguments are similar to those in [1], and we formulate the problem here so as to use the earlier results whenever possible.

2. Assumptions for Section 3

K *denotes an arbitrary real number* (independent of x,ξ,n,a) and its value may change from usage to usage. $G_{xx}(x)$ denotes the Hessian matrix of a function G and E_n^a denotes conditioning on ξ_i^a, $i < n$.

Remarks on the assumptions. In order to get rate results (i.e., limit results for $U^a(\cdot)$ or U_n^a, as $a \to 0$, $n \to \infty$) we obviously require that the tails of $\{X_n^a\}$ converge in some sense as $a \to 0$. This requires some stability properties of the "deterministic" part of (1.2), in particular that a solution x_t = constant = θ of the ODE $\dot{x} = g(x)$ is globally asymptotically stable. For notational convenience in Section 3 and in the assumptions, we set $\theta = 0$ there, without loss of generality. In Section 4 on, we reintroduce θ. It seems best to deal with the stability problem by introducing a Liapunov function $V(\cdot)$ for $\dot{x} = g(x)$. Conditions (A6)-(A7) below are often guaranteed by various forms of strong mixing conditions on $\{\xi_n^a\}$. In the usual applications to identification and adaptive systems theory [5], [8], there is an asymptotically stable A such that $g(x) = Ax$ and an affine function $f(\cdot)$ such that $f(x,\xi) = f(x)\xi$. Then $V(\cdot)$ is chosen to be a quadratic form and (A4)-(A5) hold, and so do (A6)-(A7) under simple conditions on $\{\xi_n^a\}$. See Section 7 for more detail.

A1. For each a, $\{\xi_n^a\}$ is a bounded random sequence and $Ef(x,\xi_n^a) = 0$, all x,a,n.

A2. $g(\theta) = 0$, $\theta = 0$ (here and in Section 3, for notational convenience only) $g(\cdot)$ and $f(\cdot,\cdot)$ are measurable. The first and second partial x-derivatives of $f(\cdot,\xi)$ and $g(\cdot)$ are continuous for each ξ.

A3. There is a non-negative three times continuous differentiable Liapunov function $V(\cdot)$ for $\dot{x} = g(x)$ such that $V(x) \geq 0$, $V(x) \to \infty$ as $|x| \to \infty$, $V(x) = x'Qx + o(|x|^2)$ for some positive definite matrix Q.

A4. For some real $\gamma > 0$, $V_x'(x)g(x) \leq -\gamma V(x)$.

A5. $V_{xx}(\cdot)$ is uniformly bounded and $|f(x,\xi)|^2 + |g(x)|^2 \leq K(V(x)+1)$.

A6. $\sum_{i=n}^{\infty} a|E_n^a V_x'(x)f(x,\xi_i^a)| \leq aK(V(x)+1)$.

A7. $\sum_{i=n}^{\infty} a|E_n^a(V_x'(x)f(x,\xi_i^a))_{xx}| \le aK,$

$$\sum_{i=n}^{\infty} a|E_n^a(V_x'(x)f(x,\xi_i^a))_x| \le aK(V^{1/2}(x)+1).$$

(A5) implies that f and g grow at most linearly in x.

3. Tightness of $\{U_n^a$ small a, $n \ge N_a\}$

Fix $K_0 > 0$. Let N_a denote any integer such that $\exp(-(a\gamma/2)N_a) \le K_0 a$. We have the $n \ge N_a$ requirement because of the effect of the initial condition. In general, $\{X_n/\sqrt{a}, n > 0,$ small $a\}$ will not be tight unless $X_0 = 0$. So we wait (N_a steps) until the effects of the initial condition are small. In any case, we are concerned with the tail of $\{U_n^a\}$ for small a. For the special case (1.3), it is possible to center the sequence $\{U_n^a\}$ in such a way that $\{U_n^a, n \ge 0\}$ can be dealt with (then $N_a = 0$ is used) and a better result obtained. See Section 7.

Theorem 1. *Under* (A1)-(A7), $\{U_n^a$, *small* a, $n \ge N_a\}$ *is tight*.

Proof. Again, K defines a constant whose value may change from usage to usage. Define (well-defined by (A6), recall $t_n = an$).

$$V_1^a(x,t_n) = a \sum_{i=n}^{\infty} E_n^a V_x'(x)f(x,\xi_i^a) \tag{3.1}$$

and define

$$V^a(x,t_n) = V(x) + V_1^a(x,t_n). \tag{3.2}$$

The proof uses a Liapunov function approach with Liapunov function V^a. The reason for the introduction of the V_1^a term will be clear below; basically, it is useful owing to the non-independence of the $\{\xi_i^a\}$ and allows us to "average" out their effects. Indeed, $V_1^a = 0$ when the $\{\xi_i^a\}$ are independent. We first evaluate

$$E_n^a V^a(X_{n+1}^a, t_{n+1}) - V^a(X_n^a, t_n) = T_1 + T_2 + T_3,$$

where

$$T_1 = E_n^a V(X_{n+1}^a) - V(X_n^a)$$

$$T_2 = E_n^a V_1^a(X_n^a, t_{n+1}) - V_1^a(X_n^a, t_n)$$

$$T_3 = E_n^a V_1^a(X_{n+1}^a, t_{n+1}) - E_n^a V_1^a(X_n^a, t_{n+1}).$$

Let X_n^+ and X_n^{++} denote random variables in the range $[X_n^a, X_{n+1}^a]$. Then, via truncated Taylor series expansions

$$T_1 = aV_x'(X_n^a)g(X_n^a) + aV_x'(X_n^a)f(X_n^a,\xi_n^a)$$
$$+ \frac{a^2}{2}(f(X_n^a,\xi_n^a) + g(X_n^a))'V_{xx}(X_n^+)(f(X_n^a,\xi_n^a) + g(X_n^a))$$

$$T_2 = -aV_x'(X_n^a)f(X_n^a,\xi_n^a)$$

$$T_3 = a \sum_{i=n+1}^{\infty} E_n^a V_x'(X_{n+1}^a) f(X_{n+1}^a, \xi_i^a) - a \sum_{i=n+1}^{\infty} E_n^a V_x'(X_n^a) f(X_n^a, \xi_i^a)$$

$$= a^2 \sum_{i=n+1}^{\infty} E_n^a (V_x(X_n^a) f(X_n^a, \xi_i^a))_x' (f(X_n^a, \xi_n^a) + g(X_n^a))$$

$$+ \frac{a^3}{2} \sum_{i=n+1}^{\infty} E_n^a (f(X_n^a, \xi_n^a) + g(X_n^a))' [V_x'(X_n^{++}) f(X_n^{++}, \xi_i^a)]_{xx} (f(X_n^a, \xi_n) + g(X_n^a)).$$

Now, (A4)-(A7) yield (note that T_2 cancels the second term of T_1; this is the reason for the introduction of V_1^a)

(3.3a) $$E_n^a V^a(X_{n+1}^a, t_{n+1}) - V^a(X_n^a, t_n) \le -a\gamma V(X_n^a) + a^2 K[V(X_n^a)+1],$$

By (A6), $|V_1^a(x,t_n)| \le aK(V(x)+1)$ and by (3.3a)

(3.3b) $$E_n^a V^a(X_{n+1}^a, t_{n+1}) - V^a(X_n^a, t_n) \le -a\gamma V^a(X_n^a, t_n) + a^2 K[V^a(X_n^a, t_n)+1].$$

Let $a^2 K \le a\gamma/2$ (or, equivalently $a \le a_0 \equiv \gamma/2K$). Then (3.3b) yields

(3.4) $$E_0^a V^a(X_n^a, t_n) \le \exp(-a\gamma n/2) V^a(X_0^a, 0) + Ka.$$

Equation (3.4) also holds for V replacing V^a. Thus, by (3.4) and (A3), for any constant K_1 and $n \ge N_a$, $a \le a_0$.

$$P\{X_n^{a'}QX_n^a + o(|X_n^a|^2) \geq k_1 a\} \leq \frac{K[\exp -a\gamma N_a/2][V(X_0^a)+1] + Ka}{k_1 a}$$

(3.5)

$$\leq K/k_1.$$

Tightness of $\{U_n^a$ small $a, n \geq N_a\}$ follows from (3.5) in the following way. Fix $\delta > 0$. To get tightness we need a $k_\delta < \infty$ such that $P\{\frac{X_n^{a'}QX_n^n}{a} \geq k_\delta\} \leq \delta$, all $a \leq a_0$, $n \geq N_a$. There is an $\varepsilon_0 > 0$ such that for $x'Qx \leq \varepsilon_0$, $|o(|x|^2)| \leq x'Qx/2$. For each real $k_3 > 0$, there is a $k_4(k_3) > 0$ such that $x'Qx \geq k_3$ implies $V(x) \geq k_4(k_3)$ and we can choose $k_4(\cdot)$ to be a monotonic function.

Let $n \geq N_a$. By (3.5) (recall that K might have a different value in each usage).

$$P\{X_n^{a'}QX_n^a/2a \geq k_1\} \leq K/k_1 + P\{X_n^{a'}QX_n^a \geq \varepsilon_0\}$$

$$\leq K/k_1 + P\{V(X_n^a) \geq k_4(\varepsilon_0)\} \leq K/k_1 + Ka/k_4(\varepsilon_0).$$

Choose k_1 such that $K/k_1 = \delta/2$. If $a \leq \bar{a} \equiv \delta k_4(\varepsilon_0)/2K$, then the right hand side is $\leq \delta$. If $a_0 \geq a > \bar{a}$, note that for any $k > 0$

$$P\{\frac{X_n^{a'}QX_n^a}{a} \geq k\} \leq P\{\frac{X_n^{a'}QX_n^a}{\bar{a}} \geq k\} \leq P\{V(X_n^a) \geq k_4(\bar{a}k)\}$$

$$\leq Ka/k_4(\bar{a}k) \leq Ka_0/k_4(\bar{a}k).$$

Now choose k_2 such that $Ka_0/k_4(\bar{a}k_2) \leq \delta$. Finally, let $k_\delta = \max(k_1, k_2)$. Q.E.D.

4. Remarks

(i) In a practical implementation of the algorithm (1.1) a_n might not be chosen to be constant, but might be allowed to decrease to some value $a > 0$ by iteration number N_a, where N_a will be chosen such that $E|X_n^a|^2 \simeq Ka$, and a_n will remain at value a thereafter. Under our conditions, we can, in fact, prove that $\{X_n^a/\sqrt{a_n}\}$ is tight. But if we are only interested in the "tail" of $\{X_n^a\}$, we can often assume that the initial condition error is commensurate with the value of a (i.e., $E|X_0^a|^2 \leq Ka$). We might also be more concerned with the ability of the algorithm to track changes (e.g., the changing system parameters in the identification example (Section 7), than with the transient errors. Then we only need look at the "errors" U_n^a for large n (say, $n \geq N_a$) once the transient errors due to the initial condition have been "dissipated".

(ii) Stochastic approximation (1.2) with $a_n \to 0$. Again, suppose without loss of generality that the origin is the unique asymptotically stable point of $\dot{x} = g(x)$. Let $a_n = A/(n+1)^\alpha$, $\alpha \in (0,1]$. Then the method of Theorem 1 can be used to show tightness of $\{X_n/\sqrt{a_n}, n \geq 0\}$, without the (usually required) assumption that $X_n \to 0$ w.p.1. To do this we first define $t_n = \sum_{i=0}^{n-1} a_i$ and $V_1(x,t_n) = \sum_{i=n}^{\infty} a_i E_n V_x'(x) f(x,\xi_i)$, where E_n denotes the expectation conditioned on ξ_i, $i < n$. Then under (A1)-(A5) and obvious analogs of (A6)-(A7) (the a under the summation is replaced by a_i and that on the right hand side is replaced by a_n) $\{X_n/\sqrt{a_n},\ n \geq 0\}$ can be shown to be tight. Set $V^0(x,t_n) = V(x) + V_1(x,t_n)$.

In order to prove the tightness, we derive the inequality (via the method of Theorem 1).

$$E_n V^0(X_{n+1},t_{n+1}) - V^0(X_n,t_n) \leq -\gamma a_n V(X_n) + a_n^2 K[V(X_n) + 1]$$

and then continue according to the scheme in Theorem 1 using (the analog of (3.4))

$$EV^0(X_n,t_n) \leq [\exp-\gamma t_n/2] V^0(X_0,0) + K \sum_{i=0}^{n-1} \prod_{j=i+1}^{n-1} (1-\frac{\gamma}{2} a_j + Ka_j^2) a_i^2,$$

and then show that the above right side is bounded above by Ka_n.

This result is important because the proof of tightness of $\{X_n/\sqrt{a_n}\}$ is the basic problem in rate of convergence results for stochastic approximations. If tightness of $\{X_n/\sqrt{a_n}\}$ is known, then the rate of convergence proofs in [1], [3], [9] all go through with virtually no changes without using the assumption that $X_n \to \theta \equiv 0$ w.p.1.

(iii) Stochastic approximation, additive noise. Continue with the situation in the last paragraph, but let $f(x,\xi) = \xi$, the classical Robbins-Monro case. Then (A6)-(A7) are particularly simple. There are adaptations to the Kiefer-Wolfowitz case, where $c_i = C/(i+1)^{\gamma}$, $a_i = A/(i+1)^{\alpha}$, $2\gamma < \alpha$, $\gamma > 0$, and $\{c_i\}$ is the finite difference coefficient sequence. Then the normalizing sequence is $\{\sqrt{a_n}/c_n\}$ rather than $\{\sqrt{a_n}\}$ or $\sqrt{a}$.

5. The Main Rate of Convergence Result

In this section, we let θ rather than 0 denote the stable point of $\dot{x} = g(x)$, and introduce the additional assumptions (A8)-(A12) below. Thus, we use $U_n^a = \delta X_n^a/\sqrt{a}$, where $\delta X_n^a = (X_n^a - \theta)$. For each $a > 0$, define the process $U^a(\cdot)$ by $U^a(0) = U^a_{N_a}$, and for each integer i, $U^a(t) = U^a_{i+N_a}$ in $[ia, ia+a)$. We will show that $U^a(\cdot)$ converges weakly in $D^r[0,\infty)$ to the solution to the Gauss-Markov process $U(\cdot)$:

$$(5.1) \qquad dU = \bar{H}Udt + R^{1/2}dB, \quad U(0) = \text{weak limit of } \{U^a(0)\},$$

where $H = g_x(\theta)$, $B(\theta)$ is a standard Wiener process and R is defined by (see (A9) below)

$$R = \lim_{a\to 0} \sum_{-\infty}^{\infty} R^a(i),$$

where

$$R^a(i) = Ef(\theta,\xi_j^a)f'(\theta,\xi_{j+i}^a).$$

Also, if $aN_a \to \infty$ as $a \to 0$, then the weak limit of $\{U^a(\cdot)\}$ is the stationary solution to (5.1).

Suppose that $U^a(\cdot)$ does not converge weakly to $U(\cdot)$ in $D^r[0,\infty)$. Then there is a sequence $\{a_k\}$ of positive numbers which goes to zero as fast as we wish and a $T < \infty$ such that $U^{a_k}(\cdot)$ does not converge weakly to $U(\cdot)$ in $D^r[0,T]$. Thus, it suffices to show convergence of $U^{a_k}(\cdot)$ in $D^r[0,T]$ for an arbitrary T, and for a sequence $\{a_k\}$ which goes to zero fast enough, but is otherwise arbitrary. We will set the problem up similarly to the way it was set up in [1], so that the results of that reference can be used whenever possible.

The following assumptions are required (analogous to (A3a) of [1]). Define $m_a(t) = \max\{i: ai \le t\}$. After stating the conditions, we comment on their reasonableness.

A8. <u>There is a</u> $T_1 > 0$ <u>such that</u>

$$P\{\max_{0\le t\le T_1} a\left|\sum_{i=m_a(t_N)}^{m_a(t_N+t)-1} f_x(\theta,\xi_i^a)\right| \ge \varepsilon\} \equiv \bar{k}_a(\varepsilon) \to 0$$

<u>as</u> $a \to 0$, <u>for each</u> $\varepsilon > 0$, <u>uniformly in</u> N.

A9. <u>Define</u> $h_j^a = f(\theta,\xi_j^a)$. <u>Then</u> $\{h_j^a\}$ <u>is stationary for each</u> a. <u>Define</u> $R^a(i) = Eh_j^a(h_{j+i}^a)'$. <u>Then</u> $R^a = \sum_{-\infty}^{\infty} R^a(i)$ <u>is absolutely summable and the sum converges uniformly in</u> a. <u>There is a matrix</u> R <u>such that</u> $R^a \to R$ <u>as</u> $a \to 0$.

A10. <u>Define</u> $\rho_1^a(i) = \sup_{j,\ell\ge 0} E^{1/2}|E_j^a h_{j+i}^a h_{j+i+\ell}^{a\,\prime} - R^a(\ell)|^2$. <u>Then</u> $\sum_{i=0}^{\infty} (\rho_1^a(i))^{1/2} < \infty$, <u>where the sum converges uniformly in</u> a.

A11. <u>Define</u> $\rho_2^a(i) = \sup_{k\ge 0} E^{1/2}|E_k^a h_{k+i}^a|^2$, $i \ge 0$. <u>Then</u> $\sum_{i=0}^{\infty} (\rho_2^a(i))^{1/2} < \infty$, <u>where the sum converges uniformly in</u> a.

A12. $|f_{xx}(x,\xi)| + |g_{xx}(x)| \le K$.

Remarks on (A8)-(A12). The conditions do not seem to be particularly strong. Except for the boundedness of $\{\xi_n^a\}$, they are basically the conditions used in [1], adapted to the present case.

Let $\{\xi_n^a\}$ be a ϕ-mixing process in the sense of [10] with $\sum \phi_i^{1/4} < \infty$, where ϕ_i does not depend on a. Then (A9)-(A11) hold. Condition (A3b) of [1] always holds if the noise $\{\xi_n\}$ was bounded (set $\tau = 0$ there).

Define $k_j^a = f_x(\theta,\xi_j^a)$, let there be an $\overline{R}_i$ such that $|Ek_j^a k_{j+i}^{a'}| \leq \overline{R}_i$ for all j and (small) $a > 0$, and define $\overline{R} = \sum_i \overline{R}_i$. Then by a Mensov-Rademacher type estimate ([3], p. 98), there is a K (depending on $\overline{R}$) such that for each $T_1 > 0$

$$a^2 E \max_{t \leq T_1} \Big| \sum_{i=m_a(t_N)}^{m_a(t_N+t)-1} k_i^a \Big|^2 \leq Ka^2(m_a(t_N+T_1)-m_a(t_N)) \log_2^2 4[m_a(t_N+T_1) - m_a(t_N)]$$

$$\leq T_1 K\, a \log_2^2 4T_1/a \leq K_1 a \log_2^2 a$$

which implies (A8). Other examples satisfying (A8) appear in [3].

Theorem 2. Assume (A1)-(A12). Then $\{U^a(\cdot)\}$ converges weakly in $D^r[0,\infty)$ to the $U(\cdot)$ of (5.1). If N_a is such that $aN_a \to \infty$, then $U(0)$ has the stationary distribution of $U(t)$.

Proof. Fix $T > 0$. Let $\{\varepsilon_i\}$ and a_k denote sequences of positive numbers such that $\sum_i \varepsilon_i < \infty$, $a_k \to 0$, and (see (A8))

$$\sum_k \bar{k}_{a_k}(\varepsilon_k) < \infty. \tag{5.2}$$

If (A8) holds for some T_1 then it holds for all T_1, so we can suppose that $T_1 = T$. By the discussion at the beginning of the section it is enough to prove the theorem for $\{U^{a_k}(\cdot)\}$.

Part 1. Define $\sqrt{a}\, f(\theta,\xi_j^a) \equiv \delta W_j^a$, and $W_{N,n}^a \equiv \sum_{j=N}^{N+n-1} \delta W_j^a$ and let $W^a(\cdot)$ denote the function on $[0,T]$ which equals $W_{N_a,n}^a$ on $[an,an+a)$. By a truncated Taylor series expansion

$$\begin{aligned}\delta X_{n+1}^a &= [I+ag_x(\theta)+af_x(\theta,\xi_n^a)]\,\delta X_n^a \\ &\quad + af(\theta,\xi_n^a) + aB_1(G(X_n^+),\delta X_n^a)\,\delta X_n^a \\ &\equiv [I+aH_n^a]\,\delta X_n^a + af(\theta,\xi_n^a) + a\gamma_n^a ,\end{aligned}$$

where $B_1(G(X_n^+),\delta X_n^a)$ is a matrix valued bilinear form in $G(X_n^+)$ and δX_n^a, and the elements of $G(X_n^+)$ are components of the second derivatives of $f(x,\xi_n^a) + g(x)$ evaluated at some point in the interval $[\theta,X_n^a]$. H_n^a is defined in the obvious manner. Thus

$$U_{n+1}^a = [I+aH_n^a]U_n^a + \delta W_n^a + \sqrt{a}\,\gamma_n^a. \tag{5.3}$$

Define $\Gamma^a_{N,n} \equiv \sum_{j=N}^{N+n-1} \sqrt{a}\, \gamma^a_n$ and let $\Gamma^a(\cdot)$ denote the function on $[0,T]$ which equals $\Gamma^a_{N_a,n}$ on $[an,an+a)$.

Define the function $C^\alpha_\beta(a)$ by $C^N_{N+1}(a) = I$ and for $n \geq \ell + 1$,

$$C^{N+n}_{N+\ell+1}(a) = \sum_{j=N+\ell+1}^{N+n} [I+aH^a_j] = (I+aH^a_{N+n}) \dots (I+aH^a_{N+\ell+1}).$$

By iterating (5.3) and doing a summation by parts, we get (5.4), just as (3.6) of [1] was obtained.

$$(5.4)\qquad \begin{aligned} U^a_{N+n+1} &= C^{N+n}_N(a)U^a_N + C^{N+n}_{N+1}(W^a_{N,n+1} + \Gamma^a_{N,n+1}) \\ &- \sum_{\ell=1}^{n} aC^{N+n}_{N+\ell+1}(a)H^a_{N+\ell}[(W^a_{N,n+1} - W^a_{N,\ell}) + (\Gamma^a_{N,n+1} - \Gamma^a_{N,\ell})] \end{aligned}$$

<u>Part 2</u>. We now argue that

$$(5.5)\qquad C^{m_a(t_N+t+s)}_{m_a(t_N+s)}(a_k) \to \exp \bar{H}t \quad \text{on } [0,T]$$

uniformly w.p.1, as $k \to \infty$, for <u>any fixed</u> N <u>or sequence</u> $N \to \infty$ as $k \to \infty$. The limit result (5.5) follows from [1], Lemma 2,when we make the following identification of our $\{a_k\}$ with the $\{a_k\}$ in [1], Lemma 2. To avoid confusion write the $\{a_k\}$ of [1] as $\{\bar{a}_k\}$. Then set the first $[T/a_1]$ of the $\{\bar{a}_n\}$ equal to our a_1, the next T/a_2 of the $\{\bar{a}_n\}$ equal to our a_2, etc. Then (A8), (5.2) and the Borel-Cantelli Lemma imply (A3) of [1], hence also its Lemma 2 and (5.5).

Part 3. As in [1], Theorem 2, (A9-A11) imply that $\{W^{a_k}(\cdot)\}$ converges weakly to a Wiener process $W(\cdot)$ with infinitesimal covariance R; i.e., $W(t)=R^{1/2}B(t)$, where $B(\cdot)$ is a standard Wiener process.

(5.6) $\{\Gamma^{a_k}(\cdot)\}$ converges weakly to the zero process as $k \to \infty$,

then the proof is completed via the arguments of [1], Theorem 2, part 3, and we will only prove the desired weak convergence (5.6). The purpose of the argument in [1], Theorem 2, part 3, is simply to show that the $H^a_{N+\ell}$ in (5.4) can be replaced by $g_x(\theta)$. With this replacement and the convergence of $\{U^a_N\}$ and $\{W^a(\cdot), \Gamma^a(\cdot)\}$, (5.1) follows from (5.4) and (5.5).

Let M_k denote $[T/a_k]$, and $N_k = N_{a_k}$. In view of the properties of $\{\gamma^a_i\}$, (5.6) holds if (5.7) does

$$P\{\sqrt{a_k} \sum_{i=N_k}^{N_k+M_k-1} |X^a_i|^2 \geq \varepsilon\} \to 0 \quad \text{as} \quad k \to \infty, \text{ each } \varepsilon > 0. \tag{5.7}$$

If $V(\cdot)$, the Liapunov function of Theorem 1, is quadratic, then[+] $E|X^{a_k}_{N_k+i}|^2 \leq Ka_k$ by Theorem 1, and (5.7) holds by an application of Chebychev's inequality. We now prove it in the general case.

Recall from Theorem 1 that there is a K such that for $n \geq N_a$ and the V^a of Theorem 1,

---

[+]Recall the criterion for N_{a_k} given in the first sentence of Section 3.

$$EV(X_n^a) \le Ka, \quad |EV^a(X_n^a, t_n)| \le Ka,$$

$$\text{(5.8)} \quad V^a(x, t_n) \ge -Ka,$$

$$E_n^a V(X_{n+1}^a, t_{n+1}) \le (1 - \gamma a + Ka^2) V^a(X_n^a, t_n) + a^2 K.$$

Let K be fixed at its above value henceforth in this proof, and let $n \ge N_a$ and $n - N_a \le T/a$ and let 'a' be small enough such that

$$-\gamma a + Ka^2 < 0, \quad Ka < 1, \quad a < 1.$$

Define the random variables L^a by

$$L^a(X_n^a, n) = V^a(X_n^a, t_n) + aK + (T - a(n - N_a)) a^{3/4} K.$$

Then $L^a(X_n^a, n) \ge 0$. By (5.8), we have

$$\text{(5.9a)} \quad \begin{aligned} E_n^a L^a(X_{n+1}^a, n+1) &\le (1 - \gamma a + Ka^2) V^a(X_n^a, t_n) + aK \\ &\quad + (T + aN_a - (n+1)a) Ka^{3/4} + Ka^2, \end{aligned}$$

and, consequently,

$$\text{(5.9b)} \quad \begin{aligned} &E_n^a L^a(X_{n+1}^a, n+1) - L^a(X_n^a, n) \le \\ &\le (-\gamma a + Ka^2) V^a(X_n^a, t_n) + Ka^2 - Ka^{7/4} \le 0. \end{aligned}$$

Thus $\{L^a(X_n^a, n)\}$ is a non-negative supermartingale for each small a. Thus, there is a real K_1 such that

$$(5.10)\qquad P\{\sup_{N_k\le i\le N_k+M_k-1} L^a(X_i^a,i)\ge a^{5/8}\}\le EL^a(X_{N_k}^a,N_k)/a^{5/8}$$

$$\le K(a+a^{3/4})/a^{5/8}=O(a^{1/8}).$$

There is a $K_2<\infty$ such that if $L^a(X_n^a,n)\le a^{5/8}$, then $V(X_n^a)\le K_2a^{5/8}$. We can suppose that a is small enough so that $V(x)\le K_2a^{5/8}$ implies that $V(x)\ge x'Qx/2$, and $V^a(x,n)\ge x'Qx/2-aK$. Then, for small a and $L^a(X_n^a,n)\le a^{5/8}$,

$$(5.11)\qquad 0\le L^a(X_n^a,n)=O(a^{3/4})+\delta_n,$$

where $\delta_n\ge(X_n^a)'QX_n^a/2$. Equation (5.7) follows from (5.10) and (5.11), since there is a real K_0 such that with probability $1-O(a^{1/8})$,

$$\sqrt{a}\sum_{i=N_k}^{N_k+M_k-1}(X_i^a)'QX_i^a/2\le\sqrt{a}\,(T/a)(K_0a^{5/8})=O(a^{1/8}).\quad\text{Q.E.D.}$$

6. Adaptive Systems - Examples

We will describe very briefly two of the more important systems which fall into our framework. Let $\{u_n,\mu_n\}$ and $\{y_n\}$ denote the input and output sequences, resp., of the linear system

$$(6.1)\qquad y_n=-[c_1y_{n-1}+\dots+c_ky_{n-k}]+[b_0u_n+\dots+b_\ell u_{n-\ell}]+\mu_n.$$

Suppose that the system is asymptotically stable when $u_n\equiv0$, $\mu_n\equiv0$. Define

$$\psi_n = (-y_{n-1},\dots,-y_{n-k},u_n,\dots,u_{n-\ell})', \quad \theta = (c_1,\dots,c_k,b_0,\dots,b_\ell)',$$

and let $\{\mu_n\}$ be a zero mean random sequence which is independent of the zero mean sequence $\{u_n\}$. A common algorithm for estimating θ is

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

where X_n is the n^{th} estimate of θ. Under various conditions (including $a_n \to 0$) $X_n \to \theta$ w.p.1 [2], [3]. In practice, due to extraneous noise, robustness considerations, or model uncertainties, it is common for either $a_n \downarrow a > 0$ or $a_n \equiv a$, a constant, perhaps a matrix. The case where θ varies with time and $a_n \equiv a$ is dealt with in detail in the next section.

Next, consider a similar algorithm which is very useful in adaptive communications theory. Let $\{S_{ni}\}$, $i = 1,2$ and $\{N_{ni}\}$, $i = 1,2$, represent stationary signal and noise sequences, resp. $\{S_{n1}\}$ and $\{S_{n2}\}$ ($\{N_{n1}\}$ and $\{N_{n2}\}$, resp.) are related in the sense that they are signal (noise, resp.) processes appearing at the inputs to different antennas, but are from the same transmitting source. Let $y_n = S_{n1} + N_{n1}$ and $u_n = S_{n2} + N_{n2}$ denote the actual inputs to the two antennas. Let k be a fixed integer and set $\psi_n = (u_n,\dots,u_{n-k})'$. It is desired to find the weight vector $\overline{X}$ which is the minimizing X in the expression $E[y_n - X'\psi_n]^2$. The motivation behind this desire is that (roughly speaking) if the power (in the communication theory sense) in the $\{N_{ni}\}$ sequences is greater than that in the $\{S_{ni}\}$ sequences, then under reasonable conditions the ratio of signal to noise power in the "output" difference sequence $\{y_n - \bar{X}'\psi_n\}$ is essentially the inverse of that in the input sequence $\{y_n\}$. This is obviously a desirable result. See [5] for the simple calculation, along with a discussion of some useful applications.

The algorithm (6.2) is often used to calculate the optimum X recursively, when $a_n \equiv a$. But, in this context, (6.2) is not well understood. Usually, it is only proved that EX_n converges. Exceptions to this are the work of Davisson [7] (with m-dependent stationary Gaussian sequences as inputs) and Senne [6] (where the stationary inputs satisfy a type of mixing condition), where it is proved that $E|X_n - \bar{X}|^2 \to 0$ as $a \to 0$. The method of Section 8 exploits the technique of the last section in order to get a more complete picture in the general case where a is small and the processes are non-stationary, an important case which actually justifies the use of the adaptive algorithm, but which has not yet been dealt with in the literature.

7. The Non-Stationary Identification Problem (6.1)

In this section, the parameter θ in (6.1) is allowed to vary with time, and we let θ_n^a denote its value at time n. Since the variations in $\{\theta_n^a\}$ affect the statistics of $\{\psi_n^a\}$, the identification problem is more complicated than the adaptive communications problem, and we consider only the former case. Assume that $\{u_n, \mu_n\}$ are bounded. Non-stationarities due to the θ_n^a variations are more difficult to treat than the effects of non-stationary $\{u_n,\mu_n\}$. In order to concentrate on the more important effects and minimize the notation, we assume that $\{u_n,\mu_n\}$ is stationary. Also $\{\mu_n\}$ is assumed to be zero mean, and independent of $\{u_n\}$ and $Eu_n \equiv 0$.

We now model the time variations. Let $\theta(\cdot)$ denote a uniformly continuous $R^{\ell+k+1}$ valued function on $[0,\infty)$, with values in a bounded set S. Suppose that the parameter[+] θ_n^a takes the value $\theta(an)$. To see the reasonableness of the model note that the rate of change of the θ_n^a must go to zero in some sense as $a \to 0$, for otherwise tracking would not be possible. $\theta(\cdot)$ could be a random process, but no generality is gained by that; since we treat one sample function at a time anyway. The uniform continuity condition is used to assure that the $\{y_n\}$ sequence has a certain stability property on $[0,\infty)$. We want to avoid $\theta(\cdot)$ getting "wilder and wilder" as $t \to \infty$. It is not needed if we are concerned with some "finite" interval $\{n: na \le T\}$ only.

We could allow $\{\theta_n^a\}$ to be a random sequence for each a. Even then, its rate of change must still be proportional to a in some sense (or to a fractional power of a; but then the u_n,μ_n terms play no role in the limit as $a \to 0$). In any case, we want an (limit) equation which yields the limit of the behavior of the normalized interpolation of the error $(X_n-\theta_n^a)$ process in terms of the limit of the parameter process,, so that the precise relationship can be seen. Our scheme is a natural way to get this.

[+]The parameter θ_n^a is the value of $(c_1,\ldots,c_k,b_0,\ldots,b_\ell)'$ at time n. Then the c_i,b_j are components of (hence functions of) θ_n^a at time n.

The main object is to get some information on the properties of $\{U_n^a\}$ when a is small. We might be interested, for example, in an approximation to the distribution of some continuous function of $\{U_n^a, na \leq T\}$. To get this, it makes sense to parametrize the problem so that we can get a limit result (as $a \to 0$) which will serve as the approximation to the $\{U_n^a\}$, and from which the approximation to the distributions of functions can be obtained (particularly if the convergence is in the sense of weak convergence). If we allow $a \to 0$ without simultaneously slowing down the rate of variation of θ_n^a, then obviously no limit result is possible, in general. Thus, to even discuss the behavior for small a, we must allow the θ_n^a to depend on a. As mentioned above, there are several ways in which this can be done. Our choice allows a relatively simple exhibition of the structure that the limit would have in a wide variety of cases (where, perhaps, $\theta(\cdot)$ might be a limit in some sense of the sequence of parameter variation functions $\theta^a(\cdot)$, where $\theta^a(t) = \theta_n^a$ on $[an, an+a))$.

The problem is set up in the next subsection.

The problem is formulated and some terms are defined in Subsection 7.1. Subsection 7.2 obtains estimates concerning the dependence of $y_n(\theta)$ on θ, and Subsections 7.3 and 7.4 obtain a limit theorem for the interpolation of a deterministic centering sequence $\{\bar{Y}_n\}$, and tightness of $\{U_n^a\} = \{(X_n - \theta_n^a - \bar{Y}_n)/\sqrt{a}\}$, resp. In subsection 7.5, the $C_m^n(a)$ are approximated by an exponential function and in Subsection 7.6, Theorem 5 gives the appropriate Wiener process limits and the convergence theorem for $\{U^a(\cdot)\}$.

7.1. Formulation of the problem. Let $\{y_n(\theta), \psi_n(\theta)\}$ denote the output and output-input sequence when $\theta_n^a \equiv \theta$ for all n; then $\psi_n(\theta) = \{-y_{n+1}(\theta), \ldots, -y_{n-k}(\theta), u_n, \ldots, u_{n-\ell}\}$. By (B1) below, these sequences are second order stationary for each $\theta \in S$. Define $R(\theta) = E\psi_i(\theta)\psi_i'(\theta)$, $R_t = R(\theta(t))$ and $\tilde{R}_n^a = E\psi_n\psi_n'$, the true covariance. Set $Y_n = X_n - \theta_n^a$, $\delta\theta_n^a = \theta(an+a) - \theta(an)$, $\beta_n^a = [\tilde{R}_n^a - \psi_n\psi_n']$, $\beta_n(\theta) = [R(\theta) - \psi_n(\theta)\psi_n'(\theta)]$, $\tilde{F}_i^a = E\mu_i\psi_i$, $\gamma_n^a = \mu_n\psi_n - \tilde{F}_i^a$, and $F(\theta) = E\mu_i\psi_i(\theta)$. Except for the above defined terms, the superscript "a" will normally be omitted for notational convenience, in particular on Y_n, X_n, ψ_n and $\tilde{Y}_n, \bar{Y}_n$ below. We have

$$X_{n+1} = X_n + a[(\theta_n^a)'\psi_n + \mu_n - X_n'\psi_n]\psi_n$$

$$Y_{n+1} = Y_n - \delta\theta_n^a - a\tilde{R}_n^a Y_n + a\beta_n^a Y_n + a\mu_n\psi_n.$$

Define the sequence $\{\bar{Y}_n\}$ by

(7.1a) $$\bar{Y}_{n+1} = \bar{Y}_n - \delta\theta_n^a - a\tilde{R}_n^a\bar{Y}_n + a\tilde{F}_n^a, \quad \bar{Y}_0 = Y_0,$$

and define $\tilde{Y}_n = Y_n - \bar{Y}_n$. Then

(7.1b) $$\tilde{Y}_{n+1} = \tilde{Y}_n - a\tilde{R}_n^a\tilde{Y}_n + a\beta_n^a(\tilde{Y}_n+\bar{Y}_n) + a\gamma_n^a, \quad \tilde{Y}_0 = 0.$$

$\bar{Y}_n$ is the "noiseless" part of Y_n, and contains the effects of the initial conditions. It is most convenient to work with the form $Y_n = \tilde{Y}_n + \bar{Y}_n$. This avoids the requirement of Section 3 (due to the effects of the initial condition) that $n \geq N_a$. Finally, define $\{U_n^a\} = \{\tilde{Y}_n/\sqrt{a}\}$, the sequence whose convergence we will ultimately deal with.

In order to exploit the stability properties of (6.1), it is convenient to work with (6.1) in <u>state variable form</u>. To set this up, define $\bar{u}_n = (u_n,\ldots,u_{n-\ell})'$ and $Z_n = (y_{n-k},\ldots,y_{n-1})'$. Recall that, by definition, $\theta_n^a = \theta(an)$ = value of $\{c_1,\ldots,c_k, b_0,\ldots,b_\ell\}'$ at time n, a $(k + \ell + 1)$ vector. For any S valued parameter θ, we define

$$A(\theta) = \begin{bmatrix} 0 & 1 & & 0 \\ & & \ddots & \\ & & & 1 \\ -c_k(\theta), & \ldots & \ldots, & -c_1(\theta) \end{bmatrix}, \quad B(\theta) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ b_0(\theta),\ldots,b_\ell(\theta) \end{bmatrix}, \quad C = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

$$D = [0,\ldots,0,1].$$

We define $A_i = A(\theta_i^a)$, $B_i = B(\theta_i^a)$. Then $Z_{i+1} = A_i Z_i + B_i \bar{u}_i + C\mu_i$, $y_i = DZ_{i+1}$. Define $Z_i(\theta) = \{y_{n-k}(\theta), \ldots, y_{n-1}(\theta)\}$. Then

$$(7.2) \qquad Z_{i+1}(\theta) = A(\theta)Z_i(\theta) + B(\theta)\bar{u}_i + C\mu_i, \; y_i(\theta) = DZ_{i+1}(\theta).$$

Write E_n for the expectation conditioned on $\mu_i, \bar{u}_i$, $i < n$. The following assumptions are required.

(B1) $|A^n(\theta)| \to 0$ <u>as</u> $n \to \infty$, <u>uniformly for</u> $\theta \in S$.

(B2) $\sum_{i=n}^{\infty} |E_n(\psi_i \psi_i' - \tilde{R}_i^a)| = \sum_{i=n}^{\infty} |E_n \beta_n^a|$ <u>bounded uniformly in</u> n, ω.

(B3) <u>There is a</u> $q_1 > 0$ <u>such that</u> $R(\theta) - q_1 I$ <u>is non-negative definite, for all</u> $\theta \in S$.

(B4) $\sum_{i=n}^{\infty} |E_n(\psi_i \mu_i - \tilde{F}_i^a)| = \sum_{i=n}^{\infty} |E_n \gamma_i^a|$ <u>bounded uniformly in</u> n, ω.

(B2) and (B4) are not restrictive. Under (B1), they hold under a ϕ-mixing condition (see [10] for the definition) on $\{u_n, \mu_n\}$ with $\sum \phi_i^{1/2} < \infty$. (B4) holds if the $\{\mu_i\}$ are mutually independent.

7.2. <u>Some preparatory estimates</u>. Since the statistics of the $\psi_i(\theta)$ are easier to get than those of the ψ_i, we show that ψ_i can be well approximated by $\psi_i(\theta_i^a)$, uniformly in i, for small a.

Let $P(\theta)$ denote the unique symmetric positive definite Liapunov matrix satisfying $A'(\theta)P(\theta)A(\theta) - P(\theta) = -I$. By (B1), there are $0 < \rho_1 < \rho_2 < \infty$ such that

$$\rho_1 I \leq P(\theta) \leq \rho_2 I, \text{ for all } \theta \in S. \tag{7.3}$$

We now obtain a series of results concerning the closeness of+ $R(\theta_n^a)$ to $\tilde{R}_n^a$ and $Z_{n+1}(\theta_n^a)$ to Z_{n+1}. Recall that $Z_i(\theta_n^a)$ is the value obtained from (7.2), when the θ in (7.2) is held fixed at θ_n^a for all i (i.e., $A(\theta) = A_n$, $B(\theta) = B_n$). We can write

$$Z_{n+1} = \sum_{j=-\infty}^{n} [A_n \ldots A_{j+1}](B_j\bar{u}_j + C\mu_j). \tag{7.4}$$

For small a, the sum is (7.4) converges uniformly by virtue of the stability assumption and its consequence (7.3). Indeed, by (7.3) and the fact that $|A(\theta_{n+1}^a) - A(\theta_n^a)| \to 0$ uniformly in n as $a \to 0$, there are $a_0 > 0$, $\varepsilon > 0$, $K_1 < \infty$, such that (see also [11] for a similar estimate)

$$|A_n \ldots A_{j+1}| \leq K_1(1-\varepsilon)^{n-j}, \text{ all } n,j, \text{ for } a \leq a_0. \tag{7.5a}$$

By (B1), we can suppose that K_1, ε, are chosen such that (7.5b) also holds.

$$|A_n^j| \leq K_1(1-\varepsilon)^j. \tag{7.5b}$$

The following approximation result is the basis of much of the rest of the development.

+ Recall that $R(\theta_n^a)$ is the covariance $E\psi_i(\theta_n^a)\psi_i'(\theta_n^a)$; i.e., the parameter θ is held fixed at $\theta \equiv \theta_n^a$.

<u>Lemma 1</u>. <u>Under the stability assumption (B1)</u>,

$$\sup_n |Z_{n+1} - Z_{n+1}(\theta_n^a)| \to 0 \quad \underline{\text{as}} \quad a \to 0.$$

<u>Proof</u>. Let M_a denote an integer whose specific value will be selected below. The variations in the B_j cause no problem in the proof and neither does $\{C\mu_j\}$. So, in order to simplify the proof set $B_j \equiv B$, a constant, and $C = 0$; i.e., the b_i components of $\theta(\cdot)$ are constant. Then

$$\begin{aligned}|Z_{n+1} - Z_{n+1}(\theta_n^a)| &\leq \Big|\sum_{j=-\infty}^{n-M_a} (A_n \cdots A_{n-M_a+1}) A_{n-M_a} \cdots A_{j+1} B\bar{\mu}_j\Big| \\ &+ \sum_{j=-\infty}^{n-M_a} |(A_n)^{M_a} (A_n)^{n-M_a-j} B\bar{u}_j| \\ &+ \sum_{j=n-M_a+1}^{n} |A_n \cdots A_{j+1} - A_n^{n-j}|\,|B\bar{u}_j|.\end{aligned} \tag{7.6}$$

By (7.5), there is a real K (not depending on a or M_a) such that the first two terms of (7.6) are each bounded in norm by $K(1-\varepsilon)^{M_a}/\varepsilon$. We will next get a bound on the third term. In (7.6), the value of the time parameter n plays no special role and it is enough for us to show that (7.7) tends to zero uniformly in $A_0 = A(\theta(0))$ as $a \to 0$.

$$\sum_{j=0}^{M_a-1} |A_0 \cdots A_j - A_0^{j+1}|, \tag{7.7}$$

In (7.7), A_i takes the form $A_i = A_0 + \delta_i$, and all that we assume on δ_i is that there is a real K_0 such that

$|\delta_i| \le K_0|\theta(ia) - \theta(0)|$. It is convenient to work in a matrix norm $|\cdot|_0$ which might depend on $\theta(0)$ but where there are real K_2, K_3, independent of $\theta(0)$, such that $|\cdot| \le K_3|\cdot|_0$, $|\cdot|_0 \le K_2|\cdot|$. In particular define $|A|_0^2 = \sup_{|x|=1} x'A'P(\theta(0))Ax$. Then $|A_0|_0 < 1$. By (7.3), K_2, K_3 exist and we can also suppose that $|\delta_i|_0 \le K_2|\theta(ia) - \theta(0)|$. For the j^{th} term of (7.7) in the $|\cdot|_0$ norm, we have

$$|A_0(A_0+\delta_1)\dots(A_0+\delta_j) - A_0^{j+1}|_0 \le |A_0|_0^j \sum_{i=1}^{j} |\delta_i|_0$$

$$+ |A_0|_0^{j-1} \sum_{i_2>i_1} |\delta_{i_1}\delta_{i_2}|_0 + \dots + |A_0|_0|\delta_1\dots\delta_j|_0.$$

Let $\Delta = K_2 \sup_{s\le M_a a} |\theta(s) - \theta(0)|$. Then a crude upper bound on the above is

$$|A_0|_0^{j+1}\{\binom{j}{1}(\frac{\Delta}{|A_0|_0}) + \binom{j}{2}(\frac{\Delta}{|A_0|_0})^2 + \dots + \binom{j}{j}(\frac{\Delta}{|A_0|_0})^j\}$$

$$\le |A_0|_0^{j+1}\{(1 + \frac{\Delta}{|A_0|_0})^j - 1\}.$$

and (7.7) satisfies

$$(7.8) \quad |(7.7)| \le K_3 \sum_{j=0}^{M_a-1} |A_0|_0^{j+1}\{(1 + \frac{\Delta}{|A_0|_0})^j - 1\}.$$

Now choose $M_a \to \infty$ as $a \to 0$ in such a way that aM_a (hence Δ) goes to zero. Then, since $\sup_{\theta(0)} |A(\theta(0))|_0 < 1$, the right side of (7.8) and $(1-\varepsilon)^{M_a}$ both tend to zero uniformly in $\theta(0) \in S$, as $a \to 0$. Q.E.D.

Similar proofs yield the following corollaries.

Corollary 1. Assume (B1). Let $M_a a \to 0$ and $M_a \to \infty$ as $a \to 0$ and let Θ denote the set $\{\theta(u): na - M_a a < u \leq na + M_a a\}$. Then

$$\sup_{\theta \in \Theta} |Z_{n+1}(\theta_n^a) - Z_{n+1}(\theta)| \to 0 \tag{7.9}$$

as $a \to 0$, uniformly in n.

Corollary 2. Assume (B1), (B3). Then $|R(\theta_n^a) - \tilde{R}_n^a| \to 0$ uniformly in n as $a \to 0$. Also there are $K < \infty$, $\varepsilon_0 > 0$ such that for $n > j$

$$|(I-aR(\theta_n^a))(I-aR(\theta_{n-1}^a)) \ldots (I-aR(\theta_{j+1}^a))| \leq K(1-a\varepsilon_0)^{n-j} \tag{7.10}$$

for all n,j and small a. The function $R_t = R(\theta(t))$ is continuous. The function $F(\cdot)$ is continuous and $\tilde{F}_n^a \to F(\theta(t))$, uniformly in t, as $a \to 0$, $n \to \infty$ if an is held equal to t.

Proof. The second assertion is a consequence of the continuity of $\theta(\cdot)$, and (B3) and (7.9). The rest are consequences of Lemma 1, and Corollary 1 and the details are omitted.

7.3. A limit theorem for $\{\bar{Y}_i\}$. We next turn to the treatment of the deterministic sequence $\{\bar{Y}_i\}$. Let $\Phi(t,s)$, $t > s$, denote the fundamental solution of the linear equation $\dot{x} = -R_t x$, and let $\bar{Y}^a(\cdot)$ denote the piecewise constant function on $[0,\infty)$ with values $\bar{Y}^a(t) = \bar{Y}_n$ on $[an,an+a)$, $n \geq 0$.

Lemma 2. Assume (B1), (B3). Then $\{\bar{Y}_n\}$ is uniformly bounded. If the $\tilde{R}_j^a$ in (7.12) are replaced by $R(\theta_j^a)$, then the difference between $\bar{Y}_{n+1}$ and the new right hand side converges to zero uniformly in n, as $a \to 0$. As $a \to 0$, $\bar{Y}^a(\cdot)$ converges uniformly on bounded intervals to the function $\bar{Y}(\cdot)$ defined by

$$\text{(7.11a)} \qquad \bar{Y}(t) = \Phi(t,0)\bar{Y}(0) - \int_0^t \Phi(t,s)d\theta_s + \int_0^t \Phi(t,s)F(\theta(s))ds$$

$$= \Phi(t,0)\bar{Y}(0) - \Phi(t,0)(\theta(t) - \theta(0))$$

$$- \int_0^t \Phi(t,s)R_s\,(\theta(t) - \theta(s))ds$$

$$+ \int_0^t \Phi(t,s)F(\theta(s))ds,$$

which is the unique solution to the equation

$$\text{(7.11b)} \qquad d\bar{Y}(t) = -R_t\bar{Y}(t)dt - d\theta(t) + F(\theta(t))dt.$$

Proof. For the first assertion we write the solution to (7.1a) in the form (using a summation by parts to get the second equation)

$$(7.12)\qquad \bar{Y}_{n+1} = \prod_{i=0}^{n}(I-a\tilde{R}_i^a)\bar{Y}_0$$

$$- \sum_{i=0}^{n}\prod_{j=i+1}^{n}(I-a\tilde{R}_j^a)\delta\theta_i^a + \sum_{i=0}^{n}\prod_{j=i+1}^{n}(I-a\tilde{R}_j^a)a\tilde{F}_i^a$$

$$= \prod_{i=0}^{n}(I-a\tilde{R}_i^a)\bar{Y}_0 - \prod_{i=1}^{n}(I-a\tilde{R}_i^a)\,[\theta(an+a) - \theta(0)]$$

$$- \sum_{i=1}^{n} a \prod_{j=i+1}^{n}(I-a\tilde{R}_j^a)\tilde{R}_i^a[\theta(na+a) - \theta(ia)] +$$

$$+ \sum_{i=0}^{n}\prod_{j=i+1}^{n}(I-a\tilde{R}_j^a)a\tilde{F}_i^a.$$

Now use Corollary 2 together with the boundedness of $\theta(\cdot)$.

The second assertion follows from Corollary 2. The last assertion then follows by letting $a \to 0$, $n \to \infty$, $an = t$, in (7.12) and noting that for $t > s$

$$\prod_{i=m_a(s)}^{m_a(t)}(I-aR(\theta_i^a)) \to \Phi(t,s), \quad t \geq s$$

uniformly on bounded s,t intervals. Q.E.D.

7.4. <u>Tightness of</u> $\{U_n^a\}$. With the preparatory results available, we proceed to the main result, by following the pattern of development in Theorem 1.

Theorem 3. Under (B1)-(B4), $\{U_n^a, n \geq 0, \text{small } a\}$ is tight. In particular (since $\tilde{Y}_0 = \tilde{U}_0^a = 0$), $E|\tilde{Y}_n|^2 \leq Ka$.

Proof. The proof is quite similar to that of Theorem 1 and we only remark on the basic set up. The Liapunov functions of Theorem 1 will be $V(\tilde{y}) = \tilde{y}'\tilde{y} = |\tilde{y}|^2$,

$$V_1^a(\tilde{y},t_n) = 2\tilde{y}' \sum_{i=n}^{\infty} E_n\beta_i^a\tilde{y} + 2\tilde{y}' \sum_{i=n}^{\infty} E_n\beta_i^a\bar{Y}_i + 2\tilde{y}' \sum_{i=n}^{\infty} E_n\gamma_n^a,$$

$$V^a(\tilde{y},t_n) = V(\tilde{y}) + aV_1^a(\tilde{y},t_n).$$

By virtue of (B2) and (B4), the sums are uniformly bounded and, as required by Theorem 1,

$$|V_1^a(\tilde{y},t_n)| \leq K(V(\tilde{y}) + 1). \tag{7.13}$$

Now, by applying the mechanisms of the proof of Theorem 1 and using the boundedness of $\{|\bar{Y}_i|\}$ yields

$$E_nV^a(\tilde{Y}_{n+1},t_{n+1}) - V^a(\tilde{Y}_n,t_n) \leq -a\tilde{Y}_n'\tilde{R}_n^a\tilde{Y}_n + Ka^2(1+|\tilde{Y}_n|^2). \tag{7.14}$$

Since $\tilde{R}_n^a$ is positive definite, uniformly in (small a) (Corollary 2 and (B3)), there is a $\gamma > 0$ such that $\tilde{Y}_n'\tilde{R}_n^a\tilde{Y}_n \geq \gamma V(\tilde{Y}_n)$ and the method of Theorem 1 (together with the uniform positive definiteness of $\tilde{R}_n^a$) yields the desired tightness. Q.E.D.

7.5. Approximating $C_\ell^n(a)$ by an exponential. Recall the function $C_\ell^n(a)$ introduced below (5.3). Here $-\psi_n\psi_n'$ is the H_n^a of (5.3). We can write

$$U_{n+1}^a = [I+aH_n^a]U_n^a + \sqrt{a}\ \ \beta_n^a\bar{Y}_n + \sqrt{a}\ \ \gamma_n^a. \tag{7.15}$$

The estimate (7.16) is needed in the proof of the next theorem. By (B2), (B4), (the limits of the sums are $m_a(s),m_a(t)-1$)

$$E|\sum_i \beta_i^a|^2 \le 2E\sum_i |\beta_i^a|\ |E_{i+1}\sum_{j\ge i}\beta_j^a| \le K(t-s)/a.$$

By this estimate and Chebychev's inequality there is a real K such that

$$P\{a|\sum_{i=m_a(s)}^{m_a(t)-1} \beta_i^a| \ge \varepsilon\} \le Ka(t-s)/\varepsilon^2. \tag{7.16}$$

Theorem 4. Under (B1)-(B4)

$$C_{m_a(s)}^{m_a(t)}(a) \to \Phi(t,s)$$

uniformly on bounded s,t intervals if $a \to 0$ fast enough; in particular, through any sequence $\{a_k\}$ where $\sum_k a_k < \infty$.

Proof. The proof is very similar to that of Theorem 2, Part 2. First, fix $t \le T$, let M denote an integer, and divide $[0,t]$ into M intervals, each of width δ. Suppose (without loss of generality) that $N = \delta/a$ is an integer and $\delta < 1$. The constants K below do not depend on a, δ or on $t \le T$, and their values may change from usage to usage. We have

$$(7.17) \quad |C_0^{N-1}(a) - (I + a\sum_{j=0}^{N-1} H_j^a)| \le a^2 \sum_{i_2 > i_1} |H_{i_2}^a H_{i_1}^a|$$

$$+ \dots + a^N |H_{N-1}^a \dots H_0^a| \le K\delta^2.$$

(7.17) holds (with the same K) when 0 and $N-1$ are replaced by $iN-N$ and iN, resp., for any $i > 0$. Hence,

$$|C_0^{m_a(t)}(a) - (I + a\sum_{j=NM-M}^{NM-1} H_j^a) \dots (I + a\sum_{j=0}^{N-1} H_j^a)| \le K\delta.$$

Let $\{a_k\}$ satisfy $\sum a_k < \infty$. Next, we want to show that

$$(7.19) \quad \Big| (I + a\sum_{j=NM-M}^{NM-1} H_j^a) \dots (I + a\sum_{j=0}^{N-1} H_j^a) -$$

$$(I - a\sum_{j=NM-M}^{NM-1} \tilde{R}_j^a) \dots (I - a\sum_{j=0}^{N-1} \tilde{R}_j^a)\Big| \to 0$$

uniformly for $t \in \{i\delta: i \le T/\delta\}$ w.p.1, as $a \to 0$ through the sequence $\{a_k\}$, for each fixed $\delta > 0$. Owing to the fact that both products $(I + a\sum_{i=m_a(\tau)}^{m_a(\tau+u)-1} H_j^a)$ and $C_{m_a(\tau)}^{m_a(\tau+u)}(a)$ can be made arbitrarily close to the identity by letting u and a be small, (7.19) implies that

$$|C_0^{m_a(t)}(a) - \prod_{j=0}^{m_a(t)-1} (I - a\tilde{R}_j^a)| \to 0$$

uniformly in $t \le T$, w.p.1, as $a \to 0$ through $\{a_k\}$. To get (7.19), we use the estimate

$$(7.20) \quad |(7.19)| \le Ka \sum_{i=1}^{M} \Big| \sum_{j=iN-N}^{iN-1} \beta_j^a \Big| \equiv E_0(a)$$

and, by using (7.16) with $\delta = (t-s)$ and $M = T/\delta$,

$$P\{E_0(a) \ge \epsilon\} \le \sum_{i=1}^{M} P\{a\Big|\sum_{j=iN-N}^{iN-1} \beta_j^a\Big| \ge \epsilon\delta/T\}$$

$$\le (\frac{T}{\delta})K\delta a \frac{T^2}{\epsilon^2 \delta^2} = K(\frac{T^3}{\delta^2})a.$$

Thus $\sum_k P\{E_0(a_k) \ge \epsilon\} < \infty$ and the Borel-Cantelli Lemma and (7.20) implys (7.19).

Finally, use the fact that by Corollary 2, (7.19) remains true when $\tilde{R}_j^a$ is replaced by $R(\theta(aj)) \equiv R(\theta_j^a)$ and the fact that

$$\prod_{j=0}^{NM-1} (I - aR(\theta(aj))) \to \Phi(t,0)$$

uniformly in $[0,T]$, as $a \to 0$, to complete the proof. Q.E.D.

7.6. <u>The Wiener process and the limit theorem for</u> $\{U^n(\cdot)\}$. Define $U^a(\cdot)$ as in Section 4 and define W_n^a and Γ_n^a by

$$W_n^a = \sqrt{a} \sum_{i=0}^{n-1} \beta_i^a Y_i, \quad \Gamma_n^a = \sqrt{a} \sum_{i=0}^{n-1} \gamma_i^a,$$

and let $W^a(\cdot)$ and $\Gamma^a(\cdot)$ be the continuous parameter processes with values W_n^a and Γ_n^a, resp., on $[an,an+a)$. By solving (7.15) and doing a partial summation, we get (5.4), but where U_N^a is replaced by 0 and all N's are deleted. The limit result is given in Theorem 5 under the additional assumptions:

(B5) $\{\mu_i\}$ is a sequence of bounded independent and identically distributed random variables with $E\mu_i^2 = \sigma_\mu^2$ and $E\mu_i = 0$.

(B6) $\{u_i\}$ is a bounded ϕ-mixing process [10] with mixing rate $\{\phi_i\}$ satisfying $\sum \phi_i^{1/2} < \infty$.

Remark on (B5)-(B6). They are stronger than necessary. (B5) is used because otherwise $F(\theta) \neq 0$ and it seems pointless to get a limit theorem for $U^a(\cdot)$, when the $\bar{Y}(\cdot)$ itself is biased by $F(\theta(\cdot))$. Also, (B5)-(B6) imply (B2) and also that (B4) is zero.

Theorem 5. Assume (B1), (B3), (B5), (B6). Then $\{W^n(\cdot),\Gamma^n(\cdot)\}$ converges in $D^{2(k+\ell+1)}[0,\infty)$ weakly to a Wiener process $(W(\cdot)\ \Gamma(\cdot))$, whose covariances are

$$\text{(7.22a)} \qquad \operatorname{Cov} \Gamma(t) = \sigma_\mu^2 \int_0^t R(\theta(v))dv$$

$$\text{(7.22b)} \qquad \operatorname{Cov} W(t) = \sum_{\ell=-\infty}^{\infty} \int_0^t E\beta_0(\theta(v))\bar{Y}(v)\bar{Y}'(v)\beta_\ell'(\theta(v))dv$$

$$\text{(7.22c)} \qquad E\Gamma(t)W'(t) = \sum_{\ell=1}^{\infty} \int_0^t E\mu_0\psi_0(\theta(v))\bar{Y}'(v)\beta_\ell'(\theta(v))dv.$$

$\{U^a(\cdot)\}$ converges weakly in $D^{k+\ell+1}[0,\infty)$ to the diffusion $U(\cdot)$ given by

(7.23) $$dU = -R_t U dt + dW + d\Gamma.$$

Remarks. (7.22a-c) are well defined. The sequence $\{\mu_i\}$ and, for each θ, $\{\psi_n(\theta),\beta_n(\theta)\}$, are stationary processes, so the subscript 0 and ℓ in (7.22b,c) could be i and $i+\ell$, resp., for any i. These expressions are calculated by first calculating the asymptotic moments of $\{\psi_n(\theta)\}$ needed in (7.22) for each θ. These are continuous functions of θ, so (7.22) makes sense. Note that the covariance "increment" at t depends only on the parameter $\theta(t)$, the desired form. Compare (7.22) to the R below (5.1). They are equivalent if we use $f(\theta,\xi_j^a) = \gamma_j^a + \beta_j^a\overline{Y}_j$ and neither the parameters nor $\overline{Y}_j$ vary with time.

The exact values of the covariances are complicated and one would not normally want to calculate them - even for some known "test" variation $\theta(\cdot)$. Theorem 5 gives the structure of the limit and indicates how the variances depend on the unknown function. This, in itself, is useful.

Proof. Once the assertions concerning convergence to the Wiener process are shown the proof is completed as indicated below (5.6) for Theorem 2. Only the assertions concerning the Wiener processes will be proved. The proof of those assertions are based on the proof of similar assertions in Theorem 2 and in

([1], Theorem 2). The main changes are due to the non-stationarity, which requires altering (A9)-(A11) (resp., (A6)-(A8) of [1]).

In our non-stationary and bounded $\{u_n,\mu_n\}$ case, (A10) and (A11) should be replaced by: Let $h_i^a = \gamma_i^a$ or β_i^a and define

$$\rho_1^a(i) = \sup_{j,\ell} |E_j h_{j+i}^a h_{j+i+\ell}^{a'} - Eh_{j+i}^a h_{j+i+\ell}^{a'}|, \quad \ell \geq 0,\ i \geq 0, \tag{7.24a}$$

$$\rho_2^a(i) = \sup_k |E_k h_{k+i}^a|, \quad i \geq 0. \tag{7.24b}$$

Then

$$\sum(\rho_1^a(i))^{1/2} + \sum(\rho_2^a(i))^{1/2} < \infty \tag{7.24c}$$

where the sums converge uniformly in a.

By the independence of $\{\mu_i\}$, (7.24) is obvious for $h_i^a = \gamma_i^a$. The sequence $\{\psi_i\}$ is ϕ-mixing with the corresponding $\{\phi_j\}$ satisfying $\sum \phi_i^{1/2} < \infty$, because of (B5)-(B6) and the fact that there are linear $F_n^a(\cdot)$ with uniformly (in n,q) bounded coefficients and ε_n satisfying $|\varepsilon_n^q| \leq K(1-\varepsilon)^q$ such that $y_n = F_n^q(u_n,\ldots,u_{n-q},\mu_n,\ldots,\mu_{n-q}) + \varepsilon_n^q$. Consequently, both $\{\psi_n\}$ and $\{\beta_n^a\}$ are ϕ-mixing with the corresponding $\{\phi_j\}$ satisfying $\sum \phi_j^{1/2} < \infty$. This implies (7.24) for $h_i^a = \beta_i^a$. The property (7.24) was used in ([1], Parts 1,2 of proof of Theorem 2), to show that+ $\sum_{m(t_N)}^{m(t_N+t)-1} \sqrt{a_i}\, h_i$ was tight and converged weakly to

+ In [1], $m(t) = \max\{n: \sum_0^n a_i \leq t\}$ and $a_i \to 0$ as $i \to \infty$ and $\sum a_i = \infty$; also the superscript 'a' was not used or needed. But the proof can also be used for our case, since only (7.24c) was used.

a continuous martingale, and that $\left|\sum_{m(t_n)}^{m(t_N+t)-1} \sqrt{a_i}\, h_i\right|^2$ is uniformly integrable in N. The same proof can be used when $a_i \equiv a$. Thus, $\{W^a(\cdot),\Gamma^a(\cdot)\}$ are tight in $D^{2(k+\ell+1)}[0,\infty)$ and all weak limits are continuous martingales and $\{|W^a(t)|^2, |\Gamma^a(t)|^2, \text{small } a\}$ is uniformly integrable for each t.

Choose and fix a convergent subsequence and index it by n, and let $W(\cdot),\Gamma(\cdot)$ denote the limit. As we will see, the limit will not depend on the subsequence. Let q be an arbitrary integer, and s_i, $i \le q$, t,s arbitrary except that $s_i < t < t+s$, and let $g(\cdot)$ be a bounded continuous function. Let E_t denote $E_{m_a(t)}$. By the weak convergence and uniform integrability,

$$(7.25)\qquad Eg(W^a(s_i),\Gamma^a(s_i),\ i \le q)E_t[\Gamma^a(t+s) - \Gamma^a(t)][\Gamma^a(t+s) - \Gamma^a(t)]'$$

$$\to Eg(W(s_i),\Gamma(s_i),\ i \le q)\,[\Gamma(t+s) - \Gamma(t)][\Gamma(t+s) - \Gamma(t)]'.$$

Evaluating the $E_t[\quad]$ term and using the independence of the $\{\mu_i\}$, yields (limits of the sums are $m_a(t),m_a(t+s)-1$)

$$(7.26)\qquad E_t[\Gamma^a(t+s) - \Gamma^a(t)][\Gamma^a(t+s) - \Gamma^a(t)]' = a\, E_t \sum \gamma_i^a(\gamma_i^a)'$$

$$= a \sum \sigma_\mu^2 E_t \psi_i(\psi_i)'.$$

Since $\lim |E_t\psi_i\psi_i' - \tilde{R}_i^a| \to 0$ as $|i-m_a(t)| \to \infty$ by (B5), (B6), the limit of the right side is the limit of $\sum\sigma_\mu^2\tilde{R}_i^a$, which (in turn) is the limit of $a \sum \sigma_\mu^2 R(\theta_i^a)$ which (in turn) equals $\int_t^{t+s} \sigma_\mu^2 R(\theta(v))dv$. Due to the arbitrariness of s_i,q,g,s,t, we have that

$$E\{[\Gamma(t+s) - \Gamma(t)][\Gamma(t+s) - \Gamma(t)]' \mid \Gamma(v), W(v), v \leq t\} =$$

$$= \int_t^{t+s} \sigma_\mu^2 R(\theta(v))dv,$$

hence that the right side of (7.22a) is the quadratic covariation of $\Gamma(\cdot)$. Thus $\Gamma(\cdot)$ is a Wiener process. Similarly, if the right sides of (7.22b,c) are the quadratic covariation of $W(\cdot)$ and the cross quadratic covariation of $W(\cdot),\Gamma(\cdot)$, then $(W(\cdot), \Gamma(\cdot))$ is the asserted Wiener process, and the proof will be completed.

We now do a similar calculation for $W^n(\cdot)$. We need only show that (limits of sums are $m_a(t), m_a(t+s) - 1$ unless otherwise written)

$$a E_t \sum_i \beta_i^a \bar{Y}_i \sum_j \bar{Y}_i' \beta_i^{a'} \tag{7.27}$$

converges to the integral in (7.22b) with limits $(t,t+s)$ instead of $(0,t)$. Equation (7.26) equals (use the convention $\sum_c^b = 0$ if $b < c$)

$$\sum_{\ell \geq 0} \sum_{i=m_a(t)}^{m_a(t+s)-\ell-1} a E_t \beta_i^a \bar{Y}_i \bar{Y}_{i+\ell}' \beta_{i+\ell}^{a'} \tag{7.28}$$

$$+ \sum_{\ell < 0} \sum_{i=m_a(t)+|\ell|}^{m_a(t+s)-1} a E_t \beta_i^a \bar{Y}_i \bar{Y}_{i+\ell}' \beta_{i+\ell}^{a'} .$$

For all $i, i + \ell$ in the range of the above sums, the ϕ-mixing implies that

$$|E_t \beta_i^a \bar{Y}_i \bar{Y}'_{i+\ell} \beta_{i+\ell}^a| \leq K\phi_{|\ell|}^{1/2}.$$

Since $\sum_{|\ell|\geq L} \sum_{i=m_a(t)}^{m_a(t+s)} a\phi_{|\ell|}^{1/2} \to 0$ as $L \to \infty$, we may evaluate the limit of (7.28) by evaluating the limit of the inner sums individually as $a \to 0$, and then summing over ℓ. By the same argument which we used for $\Gamma^a(\cdot)$ below (7.26), the limit of the ℓ^{th} inner sum is the same as the limit when E_t is replaced by E. Furthermore, by Lemma 1 and its Corollaries, β_i^a can be replaced by $\beta_i(\theta_i^a)$ without altering the limit. Upon making these replacements, we see that ℓ^{th} inner sum converges to the ℓ^{th} integral in (7.22b) with limits $(t,t+s)$ instead of $(0,t)$. By the argument used in connection with $\Gamma(\cdot)$, this implie that $W(\cdot)$ is a Wiener process with the asserted covariance.

We need only show that (7.22c) is the cross-quadratic covariance between $\Gamma(\cdot)$ and $W(\cdot)$ is (7.22c). The proof of this is the same as that just given for $W(\cdot)$ above. The sum is $\sum_1^{\infty}$ rather than $\sum_{-\infty}^{\infty}$, since μ_n is independent of y_i, $i < n$, and of ψ_i and β_i^a, $i \leq n$. Q.E.D.

REFERENCES

[1] H.J. Kushner, Hai Huang,"Rates of convergence for stochastic approximation type algorithms", to appear SIAM J. on Control and Optimization.

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

[3] H.J. Kushner, D.S. Clark, Stochastic Approximation Methods for Constrained and Unconstrained Systems, Appl. Math. Sci. Series, no. 26(1978), Springer-Verlag, Berlin.

[4] M.T. Wasan , Stochastic Approximation, Cambridge Univ. Press, Cambridge, 1969.

[5] B. Widrow, etal., "Stationary and nonstationary learning characteristics of the LMS adaptive filter", Proc. IEEE, 64 (1976), pp. 1151-1162.

[6] K. Senne, "Adaptive linear discrete-time estimation", Stanford Univ. Rept. SEL 68-090, June 1968.

[7] J.K. Kim, L.D. Davisson, "Adaptive linear estimation for stationary M-dependent processes", IEEE Trans. on Information Theory, IT-21, 1975, pp. 23-31.

[8] L. Ljung, "On positive real transfer functions and the convergence of some recursive schemes", IEEE Trans. on Automatic Control, AC-22, 1977, pp. 539-550.

[9] H.J. Kushner, "Rates of convergence for sequential monte-carlo optimization methods", SIAM J. on Control and Optimiz., 16(1978), pp. 150-168.

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

[11] C. Desoer, "Slowly varying system $\dot{x} = A(t)x$", IEEE Trans. on Automatic Control, AC-14(1969), p. 780.