

COLORADO STATE
UNIVERSITY
FORT COLLINS, COLORADO
80523

department of electrical engineering



STOCHASTIC APPROXIMATION WITH CORRELATED DATA

Technical Report

by

David C. Farden

AD A024279

ARPA Order Number:

1872

Contract Number:

10292750

Program Code No.:

5G10

Naval Undersea Center Contract Number:

N66001-75-C-0224

Principal Investigators:

David C. Farden
(303) 491-6018

Louis L. Scharf
(303) 491-5065

Name of Contractor:

Colorado State University

Project Engineer:

Darrell E. Marsh
(714) 225-6135

Effective Date of Contract:

30 May 1975

Short Title of Work:

Communication Receiver Studies-II

Contract Expiration Date:

1 March 1976

Amount of Contract:

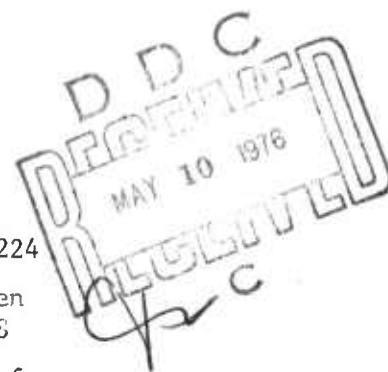
\$14,998.00

Sponsored in part by
Advanced Research Projects Agency
ARPA Order No. 1872

The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited



**BEST
AVAILABLE COPY**

ACKNOWLEDGMENT

This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Naval Undersea Center, San Diego, under Contract Numbers N66001-74-C-0035 and N66001-75-C-0224, and in part by the Office of Naval Research under Grant N00014-75-C-0518.

ACCESSION for

N7IS	White Section	☒
U S	Gaff Section	☐
UNANNOUNCED		☐

JUSTIFICATION

BY

REGISTRATION, AVAILABILITY CODES

DATE _____ AREA _____ IN FILE _____

A

Approved for public release;
Distribution Unlimited

STOCHASTIC APPROXIMATION

WITH

CORRELATED DATA¹

by

David C. Farden

Submitted to

The Annals of Statistics

April 1976

¹This work was supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Naval Undersea Center, San Diego, under Contract Numbers N66001-74-C-0035 and N66001-75-C-0224, and in part by the Office of Naval Research under Grant N00014-75-C-0518.

The author is with the Department of Electrical Engineering, Colorado State University, Fort Collins, Colorado 80523.

American Mathematical Society 1970 subject classifications. Primary 61L20, 62M10; Secondary 62J05, 60G35, 94A05.

Key words and phrases. Stochastic approximation, Robbins-Monro procedure, moment conditions, autocovariance decay rate conditions, almost sure convergence.

SUMMARY

STOCHASTIC APPROXIMATION WITH CORRELATED DATA

New almost sure convergence results are developed for a special form of the multidimensional Robbins-Monro (RM) stochastic approximation procedure. The special form treated can be viewed as a stochastic approximation to the solution $w = w_0 \in R^p$ of the linear equations $Rw = P$, where R is a $p \times p$ positive definite symmetric matrix. This special form commonly arises in adaptive signal processing applications. Essentially, previous convergence results for the RM procedure contain a common "conditional expectation condition" which is extremely difficult (if not impossible) to satisfy when the "training data" is a correlated sequence. In contrast, the new convergence results incorporate moment conditions and covariance function decay rate conditions. The ease with which these results can be applied in many cases is illustrated.

1. Introduction. Consider the set of linear equations $Rw=P$, where R is a $p \times p$ symmetric positive definite matrix, and w and P are $p \times 1$ matrices. In case R and P are unknown, and the solution, $w = w_0 = R^{-1}P$, is desired, many techniques are available for finding an estimate of w_0 . In many adaptive signal processing applications, a recursive, computationally efficient procedure for estimating w_0 is an important issue. A suitable multidimensional version of the Robbins-Monro (RM) stochastic approximation procedure (Robbins and Monro (1951)) for recursively estimating w_0 is given by

$$(1) \quad W_{n+1} = W_n + \mu_n (P_n - F_n W_n), \quad n \geq 1,$$

where $\{P_n\}$ is a sequence of random variables, $P_n \in R^p$, $\{F_n\}$ is a sequence of random $p \times p$ matrices, $\{\mu_n\}$ is a sequence of positive constants, and $W_1 \in R^p$ is arbitrary. It is assumed that P_n and F_n are functionally independent of $W_1, W_2, \dots, W_n$. It is somewhat helpful to consider w_0 to be the vector which minimizes $\xi(w) = w'Rw - 2w'P$, where $'$ denotes matrix transpose. Interpreting F_n and P_n as "instantaneous estimates" of R and P , respectively, the relationship between (1) and (deterministic) steepest descent procedures is obvious. Consequently, the family of algorithms represented by (1) has an interpretation as a family of "stochastic gradient-following" algorithms.

Algorithm (1) provides a suitable framework for the analytical treatment of many of the algorithms that have been proposed in the engineering literature for adaptive signal processing applications (e.g., see Sakrison (1966) or Farden (1975)). For such applications, any "conditional expectation assumption" is extremely difficult (if not impossible) to establish. Such conditions are commonly required by

existing convergence theorems treating the RM procedure. For example, see Schmetterer (1961, 1969), Sakrison (1966), or Farden (1975) for a discussion of existing results. The practical application of existing convergence results for the RM procedure to (1) essentially requires that $\{P_n - F_n w\}$ is an uncorrelated sequence for all fixed parameter $w \in R^P$. The special form of the RM procedure represented by (1) enables us to obtain convergence results which make use of no such conditional expectation requirements and have a decidedly different flavor than existing results for the RM procedure.

The contents and organization of this paper are as follows. Notation and basic assumptions are presented in Section 2. The framework presented in Assumption (2.1) establishes that the sequences $\{F_n\}$ and $\{P_n\}$ are such that the "time averages" of $E(F_n)$ and $E(P_n)$, respectively, are equal to R and P . The generalization resulting from these definitions is applicable to cases where $E(F_n)$ and $E(P_n)$ are periodic, such as occurs in some adaptive digital communication applications. In Section 3, it is shown that Assumptions (2.1) through (2.7) are sufficient for the a.s. convergence of W_n to w_0 . The proofs of Lemma (3.1) and Theorem (3.2) below are quite similar in spirit to the proofs of Theorem (6.1) and Theorem (6.3) of Albert and Gardner (1967), respectively. However, the seemingly less restrictive assumptions made in the present work, the simplification in proof resulting for symmetric F_n , and the basic differences in the form of algorithms treated here are offered as justification for including the results of Section 3 in this paper. Furthermore, in contrast with the

assumptions made by Albert and Gardner (1967), the form of Assumptions (2.3) - (2.7) permits us to exploit the Borel-Cantelli Lemma and the results of Serfling (1970) to prove Corollary (4.5). Corollary (4.5) provides easily verified sufficient conditions for Assumptions (2.3)-(2.7), and hence, for the a.s. convergence of W_n to w_0 . Several special cases of (1) are considered in Section 5 to illustrate the application of these results. In case F_n and P_n are strongly consistent estimates of R and P , respectively, the much simpler convergence result of Section 6 is applicable.

2. Notation and basic assumptions. The norm of a $p \times p$ matrix A , denoted by $\|A\|$, is defined here by $\|A\| = \sup_{|x|=1} |Ax|$, $x \in R^p$, where R^p denotes p -dimensional Euclidean space and $|x|$, $x \in R^p$ denotes the usual p -dimensional Euclidean norm. For A real and symmetric, $\|A\| = \max_i \{|\lambda_i(A)|\}$, where $\{\lambda_i(A)\}_{i=1}^p$ are the p eigenvalues of A . The minimum and maximum eigenvalues of a $p \times p$ matrix A are denoted by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively. The element of a $p \times p$ matrix A occurring in the i th row and j th column of A is denoted by $(A)_{i,j}$. Similarly, the i th element of $x \in R^p$ is denoted by $(x)_i$. The trace of a $p \times p$ matrix A is denoted by $\text{tr}(A) = \sum_{i=1}^p (A)_{i,i}$. The symbol 0 is used to denote the additive identity for R^1, R^p , or to denote a $p \times p$ matrix of zeros. Square brackets $[]$ are used to denote integer part. Finally, subscripted variables like v_k, n_ℓ, ℓ_i , etc. are sometimes denoted by $v(k), n(\ell), \ell(i)$, etc.

All random variables are assumed to be defined on a probability space (Ω, F, P) . All relationships between random variables are to be interpreted to hold with probability one.

It is worth emphasizing that Corollary (4.5) below establishes sufficient conditions for Assumptions (2.3)-(2.7).

(2.1) ASSUMPTION. The sequence $\{W_n\}_{n=1}^\infty$, $W_n \in R^p$, satisfies the recursion

$$(1) \quad W_{n+1} = W_n + \mu_n (P_n - F_n W_n), \quad n \geq 1,$$

where $\{P_n\}$ is a sequence of random variables, $P_n \in R^p$, $\{F_n\}$ is a sequence of real symmetric nonnegative definite ($p \times p$) random matrices, $\{\mu_n\}$ is a sequence of positive real numbers, and $W_1 \in R^p$ is arbitrary.

Define $w_0 = R^{-1}P$, where $R = \lim_{n \rightarrow \infty} n^{-1} \sum_{\ell=1}^n E(F_\ell)$, $P = \lim_{n \rightarrow \infty} n^{-1} \sum_{\ell=1}^n E(P_\ell)$. It is assumed that $E(F_\ell), E(P_\ell)$ exist, that the above limits exist, and that R is positive definite. Further, it is assumed that $n^{-1} \sum_{\ell=a+1}^{n+a} E(F_\ell)$ converges uniformly to R as $n \rightarrow \infty$ for all non-negative integers a .

Defining $V_n = W_n - w_0$ and $C_n = P_n - F_n w_0$, we have

$$(2) \quad V_{n+1} = (I - \mu_n F_n) V_n + \mu_n C_n.$$

For a sequence $\{A_i\}$ of $p \times p$ matrices, define

$$(3) \quad \prod_{i=\ell}^k A_i = \begin{cases} A_k A_{k-1} \dots A_\ell, & \text{if } k \geq \ell \\ I, & \text{if } k < \ell. \end{cases}$$

$$\text{Defining } Q_{\ell,m} = \prod_{j=\ell}^m (I - \mu_j F_j), \quad \Lambda_n = \sum_{k=1}^n Q_{k+1,n} \mu_k C_k,$$

and iterating (2), one obtains

$$(4) \quad V_{n+1} = Q_{1,n} V_1 + \Lambda_n.$$

(2.2) ASSUMPTION. The sequence $\{\mu_n\}_1^\infty$ is a nonincreasing sequence of positive constants $\mu_n = O(n^{-1})$, $0 < \lim_{n \rightarrow \infty} n\mu_n < \infty$.

(2.3) ASSUMPTION. Assumptions (2.1) and (2.2) hold and $\mu_n \|F_n\| \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$.

(2.4) ASSUMPTION. Assumption (2.1) holds and $n^{-1} \sum_{k=a+1}^{a+n} F_k \xrightarrow{a.s.} R$ as $n \rightarrow \infty$ for all positive integers a .

(2.5) ASSUMPTION. There exists a sequence of random integers $\{v_k\}$ with $1 = v_1 < v_2 < v_3 < \dots$ such that, with $p_k = v_{k+1} - v_k$ and $J_k = \{v_k, v_k+1, \dots, v_{k+1}-1\}$ we have (i) $p_k = \xi + [k^\alpha]$, for some α , $0 \leq \alpha < 1$, (ii) $p_k^{-1} \lambda_{\min}(\sum_{j \in J_k} F_j) \geq \delta > 0$, and (iii) $p_k^{-1} \lambda_{\max}(\sum_{j \in J_k} F_j) < \gamma < \infty$.

The quantities ξ , δ , and γ are all random variables that are independent of k .

(2.6) ASSUMPTION. Assumptions (2.1) and (2.2) hold and there exists a random variable $S \in R^p$ such that $S_n \xrightarrow{a.s.} S$ as $n \rightarrow \infty$.

(2.7) ASSUMPTION. Assumptions (2.1) and (2.6) hold and $\|F_n(S - S_{n-1})\| \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$.

3. Almost sure convergence of W_n to w_0 .

(3.1) LEMMA. If assumptions (2.1)-(2.5) are satisfied, then

$$||Q_{1,n}||^{a.s.} \rightarrow 0.$$

PROOF. For any positive integer n , let $K = K(n)$ be such that $n \in J_K$. Then $Q_{1,n} = Q_{v(K),n} Q_{1,v(K)-1}$, and hence $||Q_{1,n}|| \leq ||Q_{v(K),n}|| \cdot ||Q_{1,v(K)-1}||$. Defining $\Gamma_k = Q_{v(k),v(k+1)-1}$, Γ_k may be expressed as

$$\Gamma_k = I - \sum_{j \in J_k} \mu_j F_j + \sum_{q=2}^{p_k} (-1)^q \sum_{L_{k,q}} \prod_{i=1}^q \mu_{\ell(i)} F_{\ell(q-i+1)},$$

where $L_{k,q} = \{\ell_i \in J_k : \ell_1 > \ell_2 > \dots > \ell_q, i=1,2,\dots,q\}$,

so that (for $\mu_{v(k+1)-1} p_k \delta < 1$)

$$||\Gamma_k|| \leq 1 - \mu_{v(k+1)-1} p_k \delta + \sum_{q=2}^{p_k} (\mu_{v(k)} p_k \gamma)^q = 1 - \mu_{v(k+1)-1} p_k \delta + \frac{\alpha_k^{2-p_k+1}}{1 - \alpha_k},$$

where $\alpha_k = \mu_{v(k)} p_k \gamma$ and the last equality holds provided that $\alpha_k \neq 1$.

Assumptions (2.2) and (2.5) imply that there exists a random positive

integer k_0 such that $\alpha_k \leq \frac{1}{2}$ and $||\Gamma_k|| \leq 1 - \beta_k$ for all $k \geq k_0$, where

$\beta_k = \frac{1}{2} \mu_{v(k+1)-1} p_k \delta$. Similarly, for all $K \geq k_0$ and $n \in J_K$, $||Q_{v(K),n}|| \leq 3/2$.

It follows that there exists a random variable M such that for all $K \geq k_0$,

$$||Q_{1,n}|| \leq \frac{3}{2} M \prod_{k=k_0}^K (1 - \beta_k) \leq \frac{3}{2} M \exp\left(-\frac{1}{2} \delta \sum_{k=k_0}^K \mu_{v(k+1)-1} p_k\right),$$

since $1-x \leq e^{-x}$ for all real x . It is easily shown that Assumptions

(2.2) and (2.5) imply that the above summation diverges to ∞ as $K \rightarrow \infty$.

Consequently, $||Q_{1,n}||^{a.s.} \rightarrow 0$ as $n \rightarrow \infty$. $\square$

(3.2) THEOREM. If Assumptions (2.1)-(2.7) are satisfied, then

$$|V_n|^{a.s.} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

PROOF. From (2.1.4) and Lemma (3.1), it remains only to show that

$|A_n|^{a.s.} \rightarrow 0$ as $n \rightarrow \infty$. From Assumptions (2.1) and (2.6), with $S_0 = 0$ and

$Q_{n+1,n} = I$, we have

$$\begin{aligned}
 (1) \quad \Lambda_n &= \sum_{k=1}^n (Q_{k,n} - Q_{k+1,n}) S_{k-1} + S_n \\
 &= - \sum_{k=1}^n Q_{k+1,n} \mu_k^F S_{k-1} + S_n.
 \end{aligned}$$

Defining

$$B_{\ell,n} = \sum_{k=\ell}^n Q_{k+1,n} \mu_k^F (S - S_{k-1}),$$

and

$$D_n = \sum_{k=1}^n Q_{k+1,n} \mu_k^F S,$$

(1) may be expressed as $\Lambda_n = B_{1,n} - D_n + S_n$.

From Lemma 1 of Albert and Gardner (1967, p. 189), $D_n = (I - Q_{1,n})S$, so that $\Lambda_n = B_{1,n} + (S_n - S) + Q_{1,n} S$. Since $S_n \xrightarrow{a.s.} S$ and $\|Q_{1,n}\| \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$ (Assumption (2.6) and Lemma (3.1)), it remains to show that $|B_{1,n}| \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$.

Using the same notation as in the proof of Lemma (3.1), $|B_{1,n}|$ may be bounded as

$$\begin{aligned}
 |B_{1,n}| &\leq \|Q_{v(K),n}\| \cdot \sum_{k=k_0}^{K-1} \pi \sum_{i=k+1}^{K-1} \Gamma_i \sum_{j \in J_k} Q_{j+1,v(k+1)-1} \mu_j^F (S - S_{j-1}) \\
 &\quad + \|Q_{v(K),n}\| \sum_{\ell=k_0}^{K-1} \pi (1 - \beta_\ell) \sum_{k=1}^{k_0-1} \pi \sum_{i=k+1}^{k_0-1} \Gamma_i \sum_{j \in J_k} Q_{j+1,v(k+1)-1} \mu_j^F (S - S_{j-1}) \\
 &\quad + \left| \sum_{k=v(K)}^n Q_{k+1,n} \mu_k^F (S - S_{k-1}) \right|.
 \end{aligned}$$

Note that for all $j, m \in J_k$ and $k \geq k_0$ we have $\|Q_{j,m}\| \leq 1 + \sum_{q=2}^p (\mu_{v(k)}^p \gamma)^q \leq \frac{3}{2}$.

Consequently, defining $d_k = \max_{j \in J_k} |F_j(S - S_{j-1})|$, there exists a random variable M_1 such that

$$|B_{l,n}| \leq \frac{9}{4} \sum_{k=k_0}^{K-1} \prod_{i=k+1}^{K-1} (1-\beta_i) p_k \mu_{\nu(k)} d_k$$

$$+ \frac{3}{2} M_1 \prod_{i=k_0}^{K-1} (1-\beta_i) + \frac{3}{2} p_K \mu_{\nu(K)} d_K.$$

Defining $a_{n,i} = \prod_{\ell=i+1}^n (1-\beta_\ell) \beta_i$, it remains only to show that

$$\sum_{k=k_0}^{K-1} a_{K-1,k} \beta_k^{-1} p_k \mu_{\nu(k)} d_k \xrightarrow{a.s.} 0 \text{ as } K \rightarrow \infty. \text{ Clearly, for all fixed}$$

$$i \geq k_0, a_{n,i} \xrightarrow{a.s.} 0 \text{ as } n \rightarrow \infty. \text{ From Lemma 1 of Albert and Gardner (1967, p. 189), } \sum_{i=k_0}^n |a_{n,i}| = 1 - \prod_{i=k_0}^n (1-\beta_i), \text{ which converges a.s. to 1}$$

as $n \rightarrow \infty$. Consequently, by the Toeplitz Lemma (e.g., see Knopp

$$(1947, p. 75)), \lim_{K \rightarrow \infty} \sum_{k=k_0}^{K-1} a_{K-1,k} \mu_{\nu(k)} d_k p_k \beta_k^{-1} = \lim_{k \rightarrow \infty} \mu_{\nu(k)} d_k p_k \beta_k^{-1}.$$

By Assumptions (2.2) and (2.5), $\mu_{\nu(k)} p_k \beta_k^{-1} = \frac{2}{\delta} \frac{\mu_{\nu(k)}}{\mu_{\nu(k+1)} - 1}$ is bounded;

hence, $|B_{l,n}| \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$. $\square$

(3.3) COROLLARY. If $\{\mu_k\}$ satisfies Assumption (2.2) and $\|F_k\|$ is a.s. uniformly bounded (in k), then Assumptions (2.3) and (2.7) may be deleted and Theorem (3.2) remains true.

PROOF. It suffices to consider $\mu_k = k^{-1}$. Since

$$|F_n(S - S_{n-1})| \leq \|F_n\| \cdot |S - S_{n-1}|, \text{ Assumption (2.6) implies Assumption (2.7).}$$

The Borel-Cantelli Lemma and the Chebychev inequality can easily be applied to show that Assumption (2.3) is satisfied. $\square$

4. Sufficient conditions for Assumptions (2.3) - (2.7). Several auxiliary lemmas will be needed before the main result of this section, Corollary (4.5) may be stated and proved. The following lemma is a reasonably straightforward extension of Theorem A presented by

Serfling (1970). Consequently, the proof will be omitted.

(4.1) LEMMA. Let $\{x_i\}$ be a sequence of random variables, $x_i \in \mathbb{R}^p$, having finite "variances" $\sigma_i^2 = E[(x_i - E(x_i))'(x_i - E(x_i))]$. For integer $n \geq 1$ define $X_{a,n} = (x_{a+1}, \dots, x_{a+n})$, $S_{a,n} = \sum_{i=a+1}^{a+n} x_i$, and $M_{a,n} = \max \{|S_{a,1}|, \dots, |S_{a,n}|\}$. For $n \leq -1$ define $X_{a,n} = (x_{a+n+1}, \dots, x_a)$, $S_{a,n} = \sum_{i=a+n+1}^a x_i$, and $M_{a,n} = \max \{|S_{a,n}|, \dots, |S_{a,-1}|\}$. For $|n| \geq 1$ let $F_{a,n}$ denote the distribution function for $X_{a,n}$, and let $g(F_{a,n})$ be a functional depending on $F_{a,n}$. Let a_0 be an arbitrary but fixed integer and let $v \geq 2$. Suppose $g(F_{a,k}) + g(F_{a+k,l}) \leq g(F_{a,k+l})$ for all $a \geq a_0$ and $1 \leq k \leq k+l$ or $a_0 - a \leq k+l \leq k \leq -1$ such that $E(|S_{a,n}|^v) \leq g^{1/v}(F_{a,n})$ for all $a \geq a_0$ and $n \geq 1$ or $a_0 - a \leq n \leq -1$. Then $E(M_{a,n}^v) \leq (\log_2 2|n|)^v g^{1/v}(F_{a,n})$ for all $a \geq a_0$ and $n \geq 1$ or $a_0 - a \leq n \leq -1$.

(4.2) LEMMA. Let $\alpha_{k,l} = \alpha_{l,k}$ and $\rho(k,l) = \rho(l,k)$ be real-valued functions defined for all non-negative integers k, l . For $1 \leq n < m$ define

$$\begin{aligned} (1) \quad \gamma_{n,m} &= \sum_{k=n}^m \sum_{l=n}^m \alpha_{k,l} \rho(k,l) \\ &= 2 \sum_{u=1}^{m-n} \sum_{k=n}^{m-u} \alpha_{k,k+u} \rho(k,k+u) + \sum_{k=n}^m \alpha_{k,k} \rho(k,k). \end{aligned}$$

Suppose that $|\rho(k,k+u)| = O(u^{-v})$ uniformly for all positive integers k . If $\alpha_{k,l} = 1$ and $0 < v < 1$, then for large $m-n$,

$$(2) \quad |\gamma_{n,m}| = O((m-n)^{2-v}).$$

Finally, if $\alpha_{k,l} = \mu_k \mu_l$, $\mu_k = O(k^{-1})$, and $0 < v < 1$, then

$$(3) \quad |\gamma_{n,m}| = O(n^{-v/(v+2)}).$$

PROOF. Suppose that $\alpha_{k,\ell} = 1$ and $|\rho(k, k+u)| = O(u^{-\nu})$, $0 < \nu < 1$, uniformly for all positive integers k . It suffices to consider

$$|\rho(k, k+u)| = (|u|+1)^{-\nu} \text{ in which case (from (1))}$$

$$|\gamma_{n,m}| \leq 2(m-n+1) \sum_{u=0}^{m-n} (u+1)^{-\nu} \leq 2(m-n+1) (1 + \int_1^{m-n+1} x^{-\nu} dx).$$

The result (2) follows easily by evaluating the above integral.

Suppose now that $\alpha_{k,\ell} = \mu_k \mu_\ell$, $\mu_k = O(k^{-1})$, and $|\rho(k, k+u)| = O(u^{-\nu})$, $0 < \nu < 1$, uniformly for all positive integers k . For this case, it suffices to consider $\mu_k = k^{-1}$ and $|\rho(k, k+u)| = (|u|+1)^{-\nu}$. Then from (1)

$$\begin{aligned} |\gamma_{n,m}| &\leq 2 \sum_{u=1}^{m-n} (u+1)^{-\nu} \int_{n-1}^{m-u} x^{-1} (x+u)^{-1} dx + (n-1)^{-1} - m^{-1} \\ &= 2 \sum_{u=1}^{m-n} u^{-1} (u+1)^{-\nu} \beta_{n,m,u} + (n-1)^{-1} - m^{-1}, \end{aligned}$$

where $\beta_{n,m,u} = \ell n ((m-u)(n+u-1)) - \ell n (m(n-1))$. If $1 \leq u \leq \ell \leq n < m$,

then $\beta_{n,m,u} \leq \ell n (n+\ell-1) - \ell n (n-1)$. If $1 \leq \ell < u < n < m-n$, then

$\beta_{n,m,u} \leq \ell n 2$. If $n \leq u \leq m-n$, then $\beta_{n,m,u} \leq \ell n (u)$.

Consequently, for all $1 \leq \ell \leq n-2 < m-n-2$,

$$\begin{aligned} |\gamma_{n,m}| &\leq 2\ell \ell n \left(\frac{n+\ell-1}{n-1} \right) + 2 \ell n 2 \sum_{u=\ell+1}^{n-1} u^{-1-\nu} \\ &\quad + 2 \sum_{u=n}^{m-n} u^{-1-\nu} \ell n u + (n-1)^{-1} - m^{-1}. \end{aligned}$$

Letting $\ell = n^\beta$, $1 > \beta > 0$, and using the fact that $\ell n (1+x) \leq x$ for all $x \geq -1$, it follows that there exist constants c_1 , c_2 , and c_3 such that

$$|\gamma_{n,m}| \leq c_1 n^{2\beta-1} + c_2 n^{-\beta\nu} + c_3 n^{-\nu/2}.$$

Finally, if $\beta = (\nu+2)^{-1}$, then $|\gamma_{n,m}| = O(n^{-\nu/(\nu+2)})$. □

(4.3) LEMMA. Define

$$\gamma_F(i, j, l, m) = \max_{|w|=1} w^* E(F_i - E(F_i)) w w^* (F_j - E(F_j)) w w^* (F_l - E(F_l)) w w^* (F_m - E(F_m)) w,$$

where $w \in \mathbb{R}^P$. Define $p_k(a) = a + [k^\alpha]$, where a is a positive integer,

$0 \leq \alpha < 1$. Define the sequence $\{v_k(a)\}$ by $v_1(a) = 1$, $v_{k+1}(a) = v_k(a) + p_k(a)$,

$k = 1, 2, \dots$, and define $J_k(a) = \{v_k(a), v_k(a)+1, \dots, v_{k+1}(a)-1\}$. If

Assumption (2.4) holds and

$$(1) \quad \sum_{k=1}^{\infty} p_k(a) \sum_{i, j, l, m \in J_k(a)} \gamma_F(i, j, l, m) < \infty,$$

for some a , $0 \leq \alpha < 1$, and for some positive integer a , then Assumption (2.5) is satisfied.

$$\text{PROOF. Define } G_k(a) = p_k^{-1}(a) \sum_{j \in J_k(a)} F_j \text{ and } R_k(a) = p_k^{-1}(a) \sum_{j \in J_k(a)} E(F_j).$$

Let ε be given such that $0 < \varepsilon < \lambda_{\min}(R)$. Since

$$\lambda_{\min}(G_k(a)) \geq \lambda_{\min}(G_k(a) - R) + \lambda_{\min}(R) \text{ and } \lambda_{\max}(G_k(a)) \leq \lambda_{\max}(G_k(a) - R) + \lambda_{\max}(R),$$

it is sufficient to show that there exists a random sequence $\{v_k(\xi)\}$, ξ an integer-valued random variable with ξ a.s. finite such that

$$0 < \lambda_{\min}(R) - \varepsilon < \lambda_{\min}(G_k(\xi) - R) + \lambda_{\min}(R) \leq \lambda_{\max}(G_k(\xi) - R) + \lambda_{\max}(R) < \lambda_{\max}(R) + \varepsilon$$

for all k , or, equivalently, that

$$\max_{|w|=1} |w^*(G_k(\xi) - R)w| < \varepsilon \text{ for } k = 1, 2, \dots. \text{ Hence, it is also sufficient to require the stronger condition that}$$

$$\max_{|w|=1} |w^*(G_k(\xi) - R_k(\xi))w| + \max_{|w|=1} |w^*(R_k(\xi) - R)w| < \varepsilon$$

for all k . By Assumption (2.1), there exists a positive integer q_1

such that for all sequences $\{v_k(a)\}$ with $a \geq q_1$ we have

$\max_{|w|=1} |w'(R_k(a) - R)w| < \varepsilon/2$. It follows from Assumption (2.4) that for all $n=1,2,\dots$, there exists an a.s. finite random variable ξ_n^* such that

$$\max_{1 \leq k \leq n} \max_{|w|=1} |w'(G_k(\xi_n^*) - R_k(\xi_n^*))w| < \varepsilon/2.$$

Let ξ_n be the smallest such ξ_n^* so that $\xi_n \geq q_1$.

It follows that

$$P(\xi_n > a > q_1) \leq P\left(\max_{|w|=1} |w'(G_k(a) - R_k(a))w| \geq \varepsilon/2 \text{ for some } k, 1 \leq k \leq n\right)$$

$$\leq \sum_{k=1}^n P\left(\max_{|w|=1} |w'(G_k(a) - R_k(a))w| \geq \varepsilon/2\right)$$

$$\leq \sum_{k=1}^n \left(\frac{\varepsilon}{2} p_k(a)\right)^{-4} E\left(\max_{|w|=1} |w'(G_k(a) - R_k(a))w|^4\right)$$

$$= \left(\frac{2}{\varepsilon}\right)^4 \sum_{k=1}^n (p_k(a))^{-4} \sum_{i,j,\ell,m \in J_k(a)} \gamma_F(i,j,\ell,m).$$

By hypothesis the above series converges as $n \rightarrow \infty$. Applying the definition of $p_k(a)$, it follows that for all $\varepsilon_1 > 0$, there exists a positive integer $N(\varepsilon_1)$ such that for all $a \geq N(\varepsilon_1)$ we have

$\lim_{n \rightarrow \infty} P(\xi_n > a > q_1) < \varepsilon_1$. Consequently, $\xi = \sup_n \xi_n$ is a.s. finite and

Assumption (2.5) is satisfied. $\square$

(4.4) LEMMA. If there exists a real-valued function $f(k, k+u) = O(|u|^{-\beta})$, $\beta > \frac{1}{4}$, uniformly for all positive integers k such that

$$|\gamma_F(i,j,\ell,m)| \leq f^2(i,j)f^2(\ell,m) + f(i,j)f(i,\ell)f(j,m)f(\ell,m),$$

then Assumption (2.5) is satisfied.

PROOF.

It suffices to consider $f(i,j) = (|i-j|+1)^{-\beta}$.

Define the ij th element of a matrix A to be $f(i,j)$. Then

$$\sum_{i=a}^b \sum_{j=a}^b \sum_{\ell=a}^b \sum_{m=a}^b f(i,j)f(i,\ell)f(j,m)f(\ell,m) = \text{tr}(A^4) \leq (\text{tr} A^2)^2.$$

The ii th element of A^2 is given by

$$\begin{aligned} (A^2)_{i,i} &= \sum_{m=a}^{i-1} (i-m+1)^{-2\beta} + \sum_{m=i+1}^b (m-i+1)^{-2\beta} + 1 \\ &\leq c_1 (i-a+1)^{1-2\beta} + c_2 (b-i+1)^{1-2\beta}, \end{aligned}$$

for $0 < \beta < \frac{1}{2}$ and some positive constants c_1 and c_2 . Consequently, for some constant c_3 , $\text{tr}(A^2) \leq c_3 (b-a+1)^{2-2\beta}$, and hence $\text{tr}(A^4) = O((b-a)^{4-4\beta})$.

From Lemma (4.2), $(\sum_{i=a}^b \sum_{j=a}^b f^2(i,j))^2 = O((b-a)^{4-4\beta})$. Clearly, (4.3.1)

be satisfied if $\sum_{k=1}^{\infty} k^{-4\alpha} k^{-4\alpha-4\alpha\beta} < \infty$. It follows that Assumption (2.5)

is satisfied for $1 > \alpha > 1/4\beta$ if $\beta > \frac{1}{4}$. $\square$

(4.5) COROLLARY. Suppose that Assumptions (2.1), (2.2), and (2.5) are satisfied. Define $\rho_F(k, \ell) = E(F_k F_\ell) - E(F_k)E(F_\ell)$, $\rho_C(k, \ell) = E(C_k C_\ell) - E(C_k)E(C_\ell)$, and $\rho_{FC}(k, \ell, n) = E((C_k - E(C_k))F_n^2(C_\ell - E(C_\ell)))$,

where all expectations are assumed to exist. Suppose that for some $v > 0$, $u^v \max \{ |\rho_F(k, k+u)|, |\rho_C(k, k+u)|, |\rho_{FC}(k, k+u, n)| \}$ is uniformly bounded for all non-negative integers k , u , and n . Further suppose that $g_n = \sum_{k=n}^{\infty} \mu_k E(C_k)$ exists and that there exists a constant $\beta > 1$ such that

$\sum_{n=1}^{\infty} |g_n|^\beta E(|F_n|^\beta) < \infty$. Then $|V_n| \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$.

PROOF. First consider Assumption (2.4). Define $S_{a,n} = \sum_{k=a+1}^{a+n}$

$(F_k - E(F_k))w$, where $w \in R^P$. Clearly, Assumption (2.4) is satisfied iff $n^{-1} |S_{a,n}| \xrightarrow{a.s.} 0$ for all positive integers a as $n \rightarrow \infty$. Define $M_{a,n} = \max \{|S_{a,1}|, \dots, |S_{a,n}|\}$. Let $\{n_k\}$ be an increasing sequence of positive integers such that $n_k \rightarrow \infty$ as $k \rightarrow \infty$. For all $n_k \leq n \leq n_{k+1}-1$,

$$(1) \quad n^{-1} |S_{a,n}| \leq n_k^{-1} |S_{a,n(k)-1}| + n_k^{-1} M_{a+n(k)-1, n(k+1)-n(k)}.$$

To apply Lemma (4.1), define $g(F_{a,n})$ as

$$(2) \quad g(F_{a,n}) = |w|^2 \sum_{k=a+1}^{a+n} \sum_{\ell=a+1}^{a+n} \|\rho_F(k, \ell)\| = O(n^{2-\nu}), \quad 0 < \nu < 1,$$

uniformly for all positive integers a , from Lemma (4.2). It is

easily seen that $g(F_{a,n})$ satisfies the needed conditions of Lemma (4.1). Letting $n_k = k^\alpha$, $n_k^{-2} E(|S_{a,n(k)}|^2)$ is summable for all $\alpha > \nu^{-1}$.

The Chebychev inequality and Borel-Cantelli Lemma thus imply that

$$n_k^{-1} |S_{a,n(k)-1}| \xrightarrow{a.s.} 0 \text{ as } k \rightarrow \infty. \text{ Letting } \xi_k = M_{a+n(k)-1, n(k+1)-n(k)},$$

$$(3) \quad n_k^{-2} E(\xi_k^2) = O(n_k^{-\nu} (\log_2 2(n_{k+1} - n_k))^2),$$

from (2) and Lemma (4.1). Substituting $n_k = k^\alpha$ into (3), the Borel-

Cantelli Lemma and the Chebychev inequality imply that $n_k^{-1} \xi_k \xrightarrow{a.s.} 0$

as $k \rightarrow \infty$ for $\alpha > \nu^{-1}$. Consequently, from (1), Assumption (2.4) is satisfied.

Now consider Assumption (2.6). Define $S_{a,m} = \sum_{k=a+m+1}^a \mu_k (C_k - E(C_k))$,

$$M_{a,m} = \max\{|S_{a,m}|, \dots, |S_{a,-1}|\}, \text{ and } g(F_{a,m}) = \sum_{k=a+m+1}^a \sum_{\ell=a+m+1}^a \mu_k \mu_\ell |\rho_C(k, \ell)|.$$

It is easily seen that $g(F_{a,m})$ satisfies the needed conditions of

Lemma (4.1). From

Lemma (4.2), $E(|S_{a,n-a-1}|^2) = O(n^{-v/(v+2)})$ for all $a > n$. With $n_k = k^\alpha$, the Chebychev inequality and the Borel-Cantelli Lemma thus imply that

$$S - S_{n(k)-1} - g_{n(k)} = \sum_{i=n_k}^{\infty} \mu_i (C_i - E(C_i)) \xrightarrow{a.s.} 0 \text{ as } k \rightarrow \infty \text{ for } \alpha > v^{-1}(v+2).$$

Consequently, $S - S_{n(k)-1} \xrightarrow{a.s.} 0$ as $k \rightarrow \infty$. For all $n_k \leq n \leq n_{k+1}-1$,

$$(4) \quad |S - S_{n-1}| \leq |S - S_{n(k+1)-1}| + \xi_k + b_k,$$

where $\xi_k = M_{n(k+1)-1, n(k)-n(k+1)}$ and b_k is a sequence of positive constants converging to zero as $k \rightarrow \infty$. From Lemmas (4.1) and (4.2),

$$(5) \quad E(\xi_k^2) = O((\log_2 2(n_{k+1} - n_k))^2 n_k^{-v/(v+2)}).$$

Substituting $n_k = k^\alpha$ into (5), the Chebychev inequality and the Borel-Cantelli Lemma imply that $\xi_k \xrightarrow{a.s.} 0$ as $k \rightarrow \infty$ for all $\alpha > v^{-1}(v+2)$.

Consequently, Assumption (2.6) is satisfied.

Finally, consider Assumption (2.7). Define $Z_{k,n} = F_n(C_k - E(C_k))$,

$$S_{a,m} = \sum_{k=a+m+1}^a \mu_k Z_{k,a+m+1}, \text{ and } g(F_{a,m}) = \sum_{k=a+m+1}^a \sum_{\ell=a+m+1}^a \mu_k \mu_\ell |\rho_{FC}(k, \ell, a+m+1)|.$$

Proceeding as above, $E(|S_{a,n-a-1}|^2) = O(n^{-v/(v+2)})$ for all $a > n$.

With $n_k = k^\alpha$, $\alpha > v^{-1}(v+2)$, we have

$$F_{n(k)}(S - S_{n(k)-1}) - F_{n(k)}g_{n(k)} = \sum_{i=n_k}^{\infty} \mu_i Z_{i,n(k)} \xrightarrow{a.s.} 0,$$

as $k \rightarrow \infty$. Since $\sum_{n=1}^{\infty} |g_n|^\beta E(|F_n|^\beta) < \infty$, the Markov inequality and the

Borel-Cantelli Lemma imply that $F_n g_n \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$. Consequently, for all $n_k \leq n \leq n_{k+1}-1$,

$$|F_n(S - S_{n-1})| \leq |F_{n(k+1)}(S - S_{n(k+1)-1})| + \xi_k + \psi_k,$$

where $\xi_k = M_{n(k+1)-1, n(k)-n(k+1)}$ and $\psi_k \xrightarrow{a.s.} 0$ as $k \rightarrow \infty$. Since ξ_k satisfies (5), $\xi_k \xrightarrow{a.s.} 0$ as $k \rightarrow \infty$. Consequently, Assumption (2.7) is satisfied.

Finally, $\beta > 1$ and $\mu_n = O(n^{-1})$ easily provide that $\mu_n ||F_n||^{a.s.} \rightarrow 0$ by the Markov inequality and the Borel-Cantelli Lemma, and hence, Assumption (2.3) is satisfied.

Consequently, from Theorem (3.2), $|V_n| \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$. $\square$

(4.6) REMARK. An important ancillary result contained in Corollary (4.5) is that sufficient conditions for the strong consistency of the usual sample covariance function are provided. For example, let $\{x_k\}_{-\infty}^{\infty}$ be a zero mean wide-sense stationary real-valued normal random process and define $\rho_x(v) = E(x_k x_{k+v})$. Consider $F_k = F_k(u) = x_k x_{k+u}$. It is easily shown that for this case, $\rho_F(k, k+v) = |\rho_x^2(v) + \rho_x(v+u)\rho_x(v-u)|$. The proof of Corollary (4.5) shows that if $\rho_x(u) = O(u^{-v})$ for $v > 0$, then $n^{-1} \sum_{k=1}^n x_k x_{k+v} \xrightarrow{a.s.} \rho_x(v)$ as $n \rightarrow \infty$. A similar result, presented as Theorem 8B of Parzen (1961), states that $n^{-1} \sum_{k=1}^{n-|v|} x_k x_{k+|v|} \xrightarrow{a.s.} \rho_x(v)$ as $n \rightarrow \infty$ provided that there exist positive constants c, q such that $n^{-1} \sum_{u=0}^{n-1} \rho_x^2(u) \leq cn^{-q}$ for all positive integers n . The conditions of Parzen (1961) are clearly satisfied by $\rho_x(u) = O(u^{-v})$ and $v > 0$.

(4.7) REMARKS. Recall that sufficient conditions for Assumption (2.5) have been presented as Lemmas (4.3) and (4.4). Lemma (4.4) is useful for several specific choices of $\{F_n\}$, as shown below in Section 5. In case $\{||F_n||\}$ is bounded, or if $\{F_n\}$ and/or $\{P_n\}$ are deterministic, then the conditions of Corollary (4.5) are simplified, as shown below in Corollaries (4.8)-(4.11).

(4.8) COROLLARY. Let ρ_F, ρ_G , and g_n be as in Corollary (4.5). Suppose that Assumptions (2.1), (2.2), and (2.5) are satisfied, and that $\{||F_n||\}$ is bounded. If there exists a $v > 0$ such that $u^v \max \{||\rho_F(k, k+u)||, |\rho_G(k, k+u)|\}$ is uniformly bounded for all non-negative

integers k and u , and there exists a $\beta > 0$ such that $\sum_{n=1}^{\infty} |g_n|^\beta < \infty$, then $|V_n|^{a.s.} \rightarrow 0$ as $n \rightarrow \infty$.

PROOF. Simply apply Corollary (3.3) to Corollary (4.5). $\square$

(4.9) COROLLARY. Suppose that F_n is deterministic, and that Assumptions (2.1) and (2.2) are satisfied. Define $\rho_P(k, \ell) = E(P_k P_\ell) - E(P_k)E(P_\ell)$, and $g_n = \sum_{k=n}^{\infty} \mu_k (E(P_k) - F_k w_0)$. If there exists a $\nu > 0$ such that $u^\nu |\rho_P(k, k+u)|$ is uniformly bounded for all non-negative integers k and u , and if $g_n \rightarrow 0$ as $n \rightarrow \infty$ then $|V_n|^{a.s.} \rightarrow 0$ as $n \rightarrow \infty$.

PROOF. An obvious consequence of Corollary (4.5). $\square$

(4.10) COROLLARY. Suppose that $\{P_n\}$ is deterministic, and that Assumptions (2.1), (2.2), and (2.5) are satisfied. Define $\rho_{FF}(k, \ell, n) = E((F_k - E(F_k))F_n^2(F_\ell - E(F_\ell)))$, and $g_n = \sum_{k=n}^{\infty} \mu_k (P_k - E(F_k) w_0)$. If there exists a $\nu > 0$ such that $u^\nu \max \{|\rho_F(k, k+u)|, |\rho_{FF}(k, k+u, n)|\}$ is uniformly bounded for all non-negative integers k , u , and n , and there exists a $\beta > 1$ such that $\sum_{n=1}^{\infty} |g_n|^\beta E(|F_n|^\beta) < \infty$, then $|V_n|^{a.s.} \rightarrow 0$ as $n \rightarrow \infty$.

PROOF. Follows directly from Corollary (4.5). $\square$

(4.11) COROLLARY. Suppose that both $\{F_n\}$ and $\{P_n\}$ are deterministic, and that Assumptions (2.1) and (2.2) are satisfied. If $\sum_{k=1}^{\infty} \mu_k (P_k - F_k w_0)$ exists, then $|V_n| \rightarrow 0$ as $n \rightarrow \infty$.

PROOF. Trivial case of Corollary (4.5). $\square$

5. Application of Corollary 4.5.

(5.1) Special families of F_n and P_n . Let $\{X_j\}_{j=-\infty}^{\infty}$ be a sequence of R^P -valued zero-mean random variables and let $\{s_j\}_{j=-\infty}^{\infty}$ be a sequence of real-valued zero-mean random variables. Define $R_{xx}(k, \ell) = E(X_k X_\ell')$,

$P_s(k, \ell) = E(s_k X_\ell)$ and $\rho_s(k, \ell) = E(s_k s_\ell)$. Suppose that $R_{xx}(k, k+u)$, $P_s(k, k+u)$, and $\rho_s(k, k+u)$ are periodic in k with period N . Define $R = N^{-1} \sum_{k=1}^N R_{xx}(k, k)$ and $P = N^{-1} \sum_{k=1}^N P_s(k, k)$. It will become apparent in what follows that R and P satisfy Assumption (2.1).

Suppose that it is desired to choose $w \in R^P$ to minimize $\xi(w) = \sum_{k=1}^N E((s_k - w' X_k)^2)$. Such problems arise frequently in adaptive transversal filter channel equalization in digital communications.

When $N = 1$, the problem reduces to the use of jointly wide-sense stationary sequences $\{s_j\}$ and $\{X_j\}$. Assuming that R is positive definite, the desired solution, w_0 , is given by $w_0 = R^{-1}P$. Assume now that R and/or P are unknown, and that it is desired to use algorithm (2.1.1), with F_n and P_n functions of the observed time series, $\{X_j\}$ and $\{s_j\}$, to recursively estimate w_0 . Obvious candidates for F_n and P_n are

$$(1) \quad F_n = K_n^{-1} \sum_{j=n-K_n+1}^n X_j X_j',$$

and

$$(2) \quad P_n = K_n^{-1} \sum_{j=n-K_n+1}^n s_j X_j,$$

where K_n is a positive integer; e.g. $K_n = 1, K$, or n . In fact, algorithms represented here by (2.1.1) with F_n and P_n given by (1) and (2) have frequently appeared in the engineering literature for consideration in a wide range of applications.

Note that if $K_n = K$ (a constant), then $E(F_n)$, $E(P_n)$, and $E(C_n)$ are all periodic (in n) with period N . Furthermore, for any $n > 0$, $N^{-1} \sum_{k=n}^{n+N} E(C_k) = 0$. If, in addition $\{\mu_k\}$ satisfies either $\mu_k = a(\lfloor \frac{k}{N} \rfloor + b)^{-1}$ or $\mu_k = a(k+b)^{-1}$, where $a > 0$ and $b \geq 0$, then it can be

shown that $|g_n| = \left| \sum_{k=n}^{\infty} \mu_k E(C_k) \right| = O(n^{-1})$. If, for example, $E(|F_n|^2)$ is bounded (in n), then $\sum_{n=1}^{\infty} |g_n|^2 E(|F_n|^2) < \infty$, thus satisfying the conditions on $\{g_n\}$ stated in Corollaries (4.5) and (4.8)-(4.10). Of course, many other choices of $\{\mu_k\}$ are permissible. Finally, if $K_n = N$, then $E(C_k) \equiv 0$, and the conditions on $\{g_n\}$ are *a fortiori* satisfied. The remaining conditions on $\{F_n\}$ and $\{P_n\}$ stated in the preceding corollaries are quite mild "asymptotic covariance decay rate" conditions. The strongest of these decay rate conditions is that imposed on γ_F via Lemma (4.3) in order to satisfy Assumption (2.5).

Regarding the covariance decay rate conditions in the corollaries of Section 4, it may be helpful to note that

$$(3) \quad \rho_C(k, \ell) = \rho_P(k, \ell) + w'_O \rho_F(k, \ell) w_O - \rho_{PF}(k, \ell) w_O - \rho_{PF}(\ell, k) w_O,$$

where ρ_P and ρ_F are defined in Corollaries (4.9) and (4.5), respectively, and $\rho_{PF}(k, \ell) = E(P'_k F_\ell) - E(P'_k)E(F_\ell)$. Furthermore, ρ_{FC} can be expressed as

$$(4) \quad \rho_{FC}(k, \ell, n) = \rho_{PFF}(k, \ell, n) + w'_O \rho_{FF}(k, \ell, n) w_O \\ - \rho_{PFF}(k, \ell, n) w_O - \rho_{PFF}(\ell, k, n) w_O,$$

where ρ_{FF} is defined in Corollary (4.10) and $\rho_{PFF}(k, \ell, n) = E((P'_k - E(P'_k)) \cdot F_n^2 (P'_\ell - E(P'_\ell)))$, and $\rho_{PFF}(k, \ell, n) = E((P'_k - E(P'_k)) F_n^2 (F_\ell - E(F_\ell)))$. The results (3) and (4) simply express the conditions on $\{F_n\}$ and $\{C_n\}$ stated in the corollaries of Section 4 in terms of similar conditions on $\{F_n\}$ and $\{P_n\}$.

In view of the widespread consideration of algorithms fitting the framework of (2.1.1), (1), and (2), it is of interest to reduce the moment conditions on $\{F_n\}$ and $\{P_n\}$ stated in the corollaries of Section 4 to moment conditions on $\{X_j\}$ and $\{s_j\}$. In case $\{X_j\}$ and $\{s_j\}$ are normally distributed, greatly simplified conditions can be established for this family of algorithms, as shown in Section (5.2).

(5.2) The normal case. Assume the same notation and structure as in Section 5.1. Further, assume that all random variables involved are jointly normally distributed. Then all of the "covariance decay rate conditions" for $\{F_n\}$ and $\{P_n\}$ stated in the corollaries of Section 4 and Section 5.1 can be expressed in terms involving the covariance functions of elements of $\{X_j\}$ and $\{s_j\}$. In order to accomplish this, some properties of joint moments of normally distributed random variables are required.

Let $z_1, z_2, \dots, z_8$ be zero-mean jointly normally distributed with $E(z_i z_j) = \sigma(i, j)$. Using the properties of the characteristic function of $z_1, z_2, \dots, z_8$, it is straightforward to show that

$$\begin{aligned} (1) \quad & E\left(\prod_{i=1}^4 (z_{2i-1} z_{2i} - \sigma(2i-1, 2i))\right) \\ &= \sum_{l_1} \dots \sum_{l_8} \sigma(l_1, l_2) \sigma(l_3, l_4) \dots \sigma(l_7, l_8), \end{aligned}$$

where the summation is over all possible ways of combining $l_1, l_2, \dots, l_8 \in \{1, 2, \dots, 8\}$, $l_i \neq l_j$ for $i \neq j$, into four distinct unordered pairs, no pair of which is (1,2), (3,4), (5,6), or (7,8). There are sixty terms in the sum.

With F_n given by (5.1.1) and $K_n = K$, and defining $f(k, l) = \max_{1 \leq i, j \leq p}$

$|(R_{xx}(k, l))_{i,j}|$, (1) can be applied to show that the conditions of Lemma (4.4) are satisfied if for some $\nu > 1/4$,

$$u^\nu \max_{1 \leq i, j \leq p} |(R_{xx}(k, k+u))_{i,j}|$$

is uniformly bounded for all non-negative integers k and u . Similarly, if for some $\nu_1 > 0$,

$$\sup_{u=1}^v \max_{1 \leq l \leq p} \{ |(P_s(k, k+u))_l|, |\rho_s(k, k+u)| \}$$

is uniformly bounded for all non-negative integers k and u , then the decay rate conditions stated in Corollary (4.5) on ρ_F , ρ_C , and ρ_{FC} are all satisfied. Consequently, each member of the family of algorithms represented by (2.1.1) with F_n and P_n given by (5.1.1) and (5.1.2), and $K_n = K$ converge a.s. to w_0 provided that, in addition, $\{\mu_k\}$ and K are chosen as discussed in Section 5.1.

For example, let $\{x_k: k \in \{0, \pm 1, \dots\}\}$ be a real-valued zero-mean wide-sense stationary autoregressive-moving average (ARMA) normal random process. With $X_k = (x_k, x_{k-1}, \dots, x_{k-p+1})'$ and $s_k = x_{k+\alpha}$ (integer α), the required decay rates of (5.1.3) are easily established. Hence, each member of the family of algorithms represented by (2.1.1) with F_n and P_n given by (5.1.1) and (5.1.2), $K_n = K$, and $\{\mu_n\}$ satisfying Assumption (2.2) converge a.s. to w_0 .

6. A simple almost sure convergence result.

(6.1) THEOREM. Suppose that Assumption (2.1) holds and that there exist sequences $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$ of non-negative real numbers (possibly random) satisfying $\|F_n - R\| \leq a_n$ and $|F_{n0} - P_n| \leq b_n$. Furthermore, suppose that there exists a positive integer n_0 (possibly random) such that for all $n \geq n_0$, $0 < \mu_n \lambda_{\min}(R) < 1$. Then for all $n \geq n_0$,

$$(1) \quad |V_{n+1}| \leq |V_{n_0}| \cdot \prod_{k=n_0}^n (1 - \mu_k d_k) + \max_{n_0 \leq k \leq n} (b_k/d_k) \cdot (1 - \prod_{j=n_0}^n (1 - \mu_j d_j)),$$

where $d_k = \lambda_{\min}(R) - a_k$. Furthermore, if $\sum_{k=1}^n \mu_k d_k \xrightarrow{a.s.} \infty$ and $b_n d_n^{-1} \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$, then $|V_n| \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$.

PROOF. From Assumption (2.1),

$$V_{n+1} = V_n - \mu_n R V_n - \mu_n (F_n V_n + F_n w_n - P_n - R V_n),$$

so that for all $n \geq n_0$,

$$\begin{aligned} (2) \quad |V_{n+1}| &\leq (1 - \mu_n \lambda_{\min}(K)) |V_n| + \mu_n \|F_n - R\| \cdot |V_n| + \mu_n |F_n w_n - P_n| \\ &\leq (1 - \mu_n d_n) |V_n| + \mu_n b_n. \end{aligned}$$

Iterating (2), for all $n \geq n_0$,

$$(3) \quad |V_{n+1}| \leq |V_{n_0}| \prod_{k=n_0}^n (1 - \mu_k d_k) + \sum_{k=n_0}^n \prod_{j=k+1}^n (1 - \mu_j d_j) \mu_k d_k (b_k d_k^{-1}).$$

Since all terms appearing in the sum in (3) are a.s. non-negative, (1)

follows immediately from (3) with the aid of Lemma 1 of Albert and

Gardner (1967, p. 189). Furthermore, if $\sum_{k=1}^n \mu_k d_k \xrightarrow{a.s.} \infty$ and $b_n d_n^{-1} \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$, then (3) and the Toeplitz Lemma show that $|V_n| \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$. $\square$

(6.2) Remark. Theorem (6.1) is applicable to the case discussed in Section 5 with $K_k = k$.

ACKNOWLEDGMENT. The author wishes to express his sincere gratitude to Dr. L. L. Scharf for introducing the author to the problem, as well as for his continued direction throughout the course of this work. Further, the author is deeply indebted to a referee who discovered a major technical flaw in the original manuscript.

REFERENCES

- [1] Albert, A. E. and Gardner, L. A., Jr. (1967). *Stochastic Approximation and Nonlinear Regression*. M. I. T. Press, Cambridge, Massachusetts.
- [2] Farden, D. C. (1975). *Stochastic Approximation with Correlated Data*. Ph.D. Dissertation, Department of Electrical Engineering, Colorado State Univ.
- [3] Knopp, K. (1947). *Theory and Applications of Infinite Series*. Hafner, New York.
- [4] Parzen, E. (1961). An Approach to Time Series Analysis. *Ann. Math. Statist.* 32 951-989.
- [5] Robbins, H., and Monro, S. (1951). A Stochastic Approximation Method. *Ann. Math. Statist.* 22 400-407.
- [6] Sakrison, D. J. (1966). Stochastic Approximation: A Recursive Method for Solving Regression Problems. In *Advances in Comm. Syst.* 2 (A. V. Balakrishnan, ed.). 51-106. Academic Press, New York and London.
- [7] Schmetterer, L. (1961). Stochastic Approximation. *Fourth Berkeley Symp. Math. Statist. Prob.* 1 587-609.
- [8] Schmetterer, L. (1969). Multidimensional Stochastic Approximation. In *Multivariate Analysis II* (P. R. Krishnaiah, ed.) 443-460. Academic Press, New York and London.
- [9] Serfling, R. J. (1970). Moment Inequalities for the Maximum Cumulative Sum. *Ann. Math. Statist.* 41 1227-1234.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Stochastic Approximation with Correlated Data.		5. TYPE OF REPORT & PERIOD COVERED Technical Report
6. AUTHOR(s) David C. Farden Louis L. Scharf		7. PERFORMING ORG. REPORT NUMBER N66001-74-C-0035 N66001-75-C-0224
8. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Electrical Engineering Colorado State University Fort Collins, Colorado 80523		9. CONTRACT OR GRANT NUMBER(s) N66001-74-C-0035 N66001-75-C-0224
10. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, VA 22209		11. PROGRAM ELEMENT PROJECT TASK AREA & WORK UNIT NUMBERS ARPA Order-1872
12. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Naval Undersea Center San Diego, California 92132		13. REPORT DATE Apr 1976
14. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		15. NUMBER OF PAGES 22
15. SECURITY CLASS. (of this report) Unclassified		16. DECLASSIFICATION DOWNGRADING SCHEDULE
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Stochastic approximation, Robbins-Monro procedure, almost sure convergence, moment conditions, covariance decay-rate conditions, correlated data. $W = W \text{ SUBVERSION } R^P$		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) New almost sure convergence results are developed for a special form of the multidimensional Robbins-Monro (RM) stochastic approximation procedure. The special form treated can be viewed as a stochastic approximation to the solution $w = w \in R^P$ of the linear equations $Rw = P$, where R is a $p \times p$ positive definite symmetric matrix. This special form commonly arises in adaptive signal processing applications. Essentially, previous convergence results for the RM procedure contain a common "conditional expectation condition" which is		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

406 434

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20. con't.

extremely difficult (if not impossible) to satisfy when the training data is a correlated sequence. In contrast, the new convergence results incorporate moment conditions and covariance function decay rate conditions. The ease with which these results can be applied in many cases is illustrated.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)