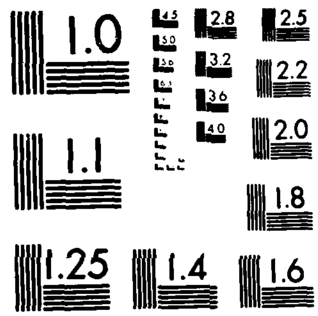
AD-A129 214 WEAK CONVERGENCE AND ASYMPTOTIC PROPERTIES OF ADAPTIVE FILTERS WITH CONST.'.(U) BROWN UNIV PROVIDENCE RI LEFSCHETZ CENTER FOR DYNAMICAL SYSTE.. 1/1

UNCLASSIFIED H J KUSHNER ET AL. 06 MAR 83 LCDS-83-7 F/G 12/1 NL

END
DATE
FILMED
7 83
DTIC



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

Lefschetz Center for Dynamical Systems

Division of Applied Mathematics Brown University Providence RI 02912

WEAK CONVERGENCE AND ASYMPTOTIC PROPERTIES
OF ADAPTIVE FILTERS WITH CONSTANT GAINS

by

Harold J. Kushner and Adam Shwartz

March 6, 1983 LCDS Report #83-7

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 83-0478	2. GOVT ACCESSION NO. AD-A129214	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) WEAK CONVERGENCE AND ASYMPTOTIC PROPERTIES OF ADAPTIVE FILTERS WITH CONSTANT GAINS		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
		6. PERFORMING ORG. REPORT NUMBER LCDS Report #83-7
7. AUTHOR(s) Harold J. Kushner and Adam Shwartz		8. CONTRACT OR GRANT NUMBER(s) AFOSR-81-0116
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lefschetz Center for Dynamical Systems, Division of Applied Mathematics, Brown University, Providence RI 02912		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE61102F; 2304/A4
11. CONTROLLING OFFICE NAME AND ADDRESS Mathematical & Information Sciences Directorate Air Force Office of Scientific Research Bolling AFB DC 20332		12. REPORT DATE MAR 83
		13. NUMBER OF PAGES 21
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

18. SUPPLEMENTARY NOTES

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

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

SEE REVERSE

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

83 06 10 034

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ITEM #20, CONTINUED:

The basic adaptive filtering algorithm $X^{\varepsilon}_{n+1} = X^{\varepsilon}_{n} - \varepsilon Y_n(Y'_n X^{\varepsilon}_n - \psi_n)$ is analyzed , via the theory of weak convergence. Apart from some very special cases, the analysis is hard when done for each fixed $\varepsilon > 0$. But the weak convergence techniques are set up to provide much information for small ε. The relevant facts from the theory are given. Define $x^{\varepsilon}(\cdot)$ by $x^{\varepsilon}(t) = X^{\varepsilon}_n$ on $[n\varepsilon, n\varepsilon + \varepsilon)$. Then weak (distributional) convergence of $\{x^{\varepsilon}(\cdot)\}$ and of $\{x^{\varepsilon}(\cdot + t_{\varepsilon})\}$ is proved under very weak assumptions, where $t_{\varepsilon} \to \infty$ as $\varepsilon \to 0$. The normalized errors $\left\{\frac{X^{\varepsilon}_n - \theta}{\sqrt{\varepsilon}}\right\}$ are analyzed, where θ is a 'stable' point for the 'mean' algorithm. We also develop the asymptotic properties of a projection algorithm, where the X^{ε}_n are truncated at each iteration, if they fall outside of a given set.

WEAK CONVERGENCE AND ASYMPTOTIC PROPERTIES
OF ADAPTIVE FILTERS WITH CONSTANT GAINS

by

Harold J. Kushner[†]

Adam Shwartz[††]

Brown University, Division of Applied Mathematics
Lefschetz Center for Dynamical Systems

March 6, 1983

[†] Division of Applied Mathematics and Engineering. This work was supported by the Air Force Office of Scientific Research under contract #AFOSR 81-0116, in part by the National Science Foundation under contract #ECS 82-11476, and in part by the Office of Naval Research under contract #N00014-76-C-0279-P0007.

[††] Division of Applied Mathematics. This work was supported by the Office of Naval Research under contract #N00014-76-C-0279-P0007.

ABSTRACT

The basic adaptive filtering algorithm $X_{n+1}^{\varepsilon} = X_n^{\varepsilon} - \varepsilon Y_n(Y_n' X_n^{\varepsilon} - \psi_n)$ is analyzed , via the theory of weak convergence. Apart from some very special cases, the analysis is hard when done for each fixed $\varepsilon > 0$. But the weak convergence techniques are set up to provide much information for small ε. The relevant facts from the theory are given. Define $x^{\varepsilon}(\cdot)$ by $x^{\varepsilon}(t) = X_n^{\varepsilon}$ on $[n\varepsilon, n\varepsilon + \varepsilon)$. Then weak (distributional) convergence of $\{x^{\varepsilon}(\cdot)\}$ and of $\{x^{\varepsilon}(\cdot + t_{\varepsilon})\}$ is proved under very weak assumptions, where $t_{\varepsilon} \to \infty$ as $\varepsilon \to 0$. The normalized errors $\left\{\frac{X_n^{\varepsilon} - \theta}{\sqrt{\varepsilon}}\right\}$ are analyzed, where θ is a 'stable' point for the 'mean' algorithm. We also develop the asymptotic properties of a projection algorithm, where the X_n^{ε} are truncated at each iteration, if they fall outside of a given set.

1. Introduction.

This paper illustrates the power of weak convergence methods through the analysis of the basic algorithm of adaptive filtering.

$$X_{n+1}^{\varepsilon} = X_n^{\varepsilon} - \varepsilon Y_n(Y_n' X_n^{\varepsilon} - \psi_n), \; X_n^{\varepsilon} \in R^r, \text{ Euclidean r-space.} \tag{1.1}$$

{Y sub n, psi sub n}

Except for the simplest cases (e.g., when $\{Y_n, \psi_n\}$ are mutually independent), the analysis of the algorith ~~(1.1)~~ for fixed epsilon $\cancel{\varepsilon}$ is difficult. However, asymptotic analysis $(\varepsilon \to 0)$ via weak convergence methods provides much information, relatively painlessly. Define the interpolated process $x^{\varepsilon}(\cdot)$ by

(epsilon approaches limit of 0)

$$x^{\varepsilon}(t) = X_n^{\varepsilon} \quad \text{on} \quad [n\varepsilon, n\varepsilon + \varepsilon). \tag{1.2}$$

Let $\{t_{\varepsilon}\}$ denote a sequence which goes to ∞ as $\varepsilon \to 0$.

In the next section, we review some definitions and results concerning weak convergence. In Section 3, the weak convergence of $\{x^{\varepsilon}(\cdot)\}$ is studied, and in Section 4, we examine the limit problem for $\{x^{\varepsilon}(t_{\varepsilon} + \cdot)\}$ to get a clearer picture of the asymptotic behavior. In Section 5, the normalized errors are studied, and a very useful projection or truncation algorithm is dealt with in Section 7.

2. Background and Definitions.

Weak Convergence. Weak convergence is an extension of the notion of convergence in distribution to random variables with values in a function space. The sequence $\{x^{\varepsilon}(\cdot)\}$ of (1.2) can be viewed as a sequence of random variables with paths in the function space $D^r[0,\infty)$, the space of R^r-valued functions on $[0,\infty)$ which are right continuous and have left hand limits. As will be seen, this point of view is very useful in applications. This space is discussed in Billingsley [1] and Kurtz [2], two excellent references for weak convergence theory. Under the so called "Skorohod topology" ([1], Section 14), $D^r[0,\infty)$ is separable and metrizable, and the metric is complete.

The space $D^r[0,\infty)$ is useful for two main reasons. First, processes with paths in $D^r[0,\infty)$ arise naturally in applications (e.g., the $x^{\varepsilon}(\cdot)$ of (1.2)). Second, its' topology is weaker than that of $C^r[0,\infty)$, the space of R^r-valued continuous functions on $[0,\infty)$, so that the criteria for compactness are less stringent, and better convergence theorems can be obtained, even if the paths or their limits lie in $C^r[0,\infty)$.

For R^r-valued random variables $\{X_n\}$, we say that $\{X_n\}$ converges weakly (or in distribution) to X iff

$$Ef(X_n) \to Ef(X) \tag{2.1}$$

for each bounded, real valued and continuous function $f(\cdot)$. If X_n and X take values in a metric space, we also say that $\{X_n\}$ converges weakly to X and write $X_n \Rightarrow X$ if (2.1) holds. When the random variables X_n are functions, we write lower case $x^n(\cdot)$. Since $f(x(\cdot)) \equiv \{x(t_1), \ldots, x(t_\ell)\}$ is a continuous function from $C^r[0,\infty)$ to $R^{r\ell}$, weak convergence $x^n(\cdot) \Rightarrow x(\cdot)$ in $C^r[0,\infty)$ implies the convergence of multivariate distributions $\{x^n(t_i), i \leq \ell\} \Rightarrow \{x(t_i), i \leq \ell\}$. Similarly on $D^r[0,\infty)$, if the limit process $x(\cdot)$ is continuous w.p.1 at $t_1, \ldots t_\ell$.

For R^r-valued X_n, the Helly-Bray Theorem states the following: If for each $\delta > 0$, $\exists$ K_δ compact such that $P\{X_n \in K_\delta\} \geq 1 - \delta$ for all n, then $\{X_n\}$ is said to be tight and it has a subsequence which converges in distribution. The definition of tightness carries over to metric space valued $\{X_n\}$. Prohorov's Theorem [1] states that tightness of $\{x^n(\cdot)\}$ implies that it has a weakly convergent subsequence.

In the sequel, we frequently use the above 'subsequence' result in the following way. First tightness is proved. Then a weakly convergent subsequence is extracted. The limit of this subsequence is then characterized as the solution to a specific ODE (ordinary differential equation) or SDE (Itô equation). It is then shown that the limit process $x(\cdot)$ does not depend on the subsequence. Hence $x^n \Rightarrow x(\cdot)$.

Let $\{X_n^\varepsilon\}$ be defined by $X_{n+1}^\varepsilon = X_n^\varepsilon + \varepsilon F_{\varepsilon,n}$, where $\{F_{\varepsilon,n}, \varepsilon > 0, n < \infty\}$ is uniformly integrable, and define $x^\varepsilon(\cdot)$ as in (1.2). Then $\{x^\varepsilon(\cdot)\}$ is tight in $D^r[0,\infty)$ and all (weak) limit processes have continuous paths. The assertion follows from ([1], Theorem 15.2).

Let $x^n(\cdot) \Rightarrow x(\cdot)$, where the paths are in $C^r[0,\infty)$ or $D^r[0,\infty)$. Since we are concerned with weak convergence only, the probability space is not important, and we can select it in any convenient way, provided only that the distributions of each $x^n(\cdot)$ and of $x(\cdot)$ does not change. By a technique known as Skorohod imbedding ([3], Theorem 3.1.1) one can choose the probability space such that $x^n(\cdot) \to x(\cdot)$ w.p.1 in the topology of the (path) space C^r or D^r. This very useful method will often be used without explicit mention.

The Martingale problem. In this paper, all limits will satisfy ODE's or SDE's. From the point of view of usefulness in weak convergence analysis, a very nice way of characterizing a limit process $x(\cdot)$ is to show that it satisfies a martingale problem (the martingale problem of Stroock and Varadhan [4]). Let $\hat{C}_0^2$ denote the space of real valued functions on R^r with compact support and continuous second partial derivatives.

Define the operator $\mathcal{L}$ on $\hat{C}_0^2$ by

$$\mathcal{L}f(x) = f_x'(x)b(x) + \frac{1}{2}\sum_{i,j} a_{ij}(x)f_{x_ix_j}(x), \tag{2.2}$$

where $\{a_{ij}(x)\} = a(x) = \sigma(x)\sigma'(x)$ and $b(\cdot)$ and $\sigma(\cdot)$ are continuous. Let $x(\cdot)$ be a random process with paths in $D^r[0,\infty)$. Define

$$M_f(t) = f(x(t)) - \int_0^t \mathcal{L}f(x(u))du.$$

If $M_f(\cdot)$ is a martingale for each $f(\cdot) \in \hat{C}_0^2$, then $x(\cdot)$ is continuous and is said to satisfy the martingale problem for operator $\mathcal{L}$. It can be shown that there is a Wiener process $w(\cdot)$ such that $x(\cdot)$ satisfies the Itô equation

$$(2.3) \qquad dx = b(x)dt + \sigma(x)dw$$

Let $x(\cdot)$ be a right continuous process and $\mathcal{T}$ a countable set. A useful method to show that $x(\cdot)$ solves the martingale problem is to show that for arbitrary $f \in \hat{C}_0^2$, arbitrary ℓ, and arbitrary $t_1 < \ldots < t_\ell < t < t + s$ <u>not</u> in $\mathcal{T}$ and arbitrary $h(\cdot)$ bounded and continuous,

$$(2.4) \qquad E[h(x(t_i), i \le \ell)[f(x(t + s)) - f(x(t)) - \int_t^{t+s} \mathcal{L} f(x(u))du]]=0$$

This implies that $x(\cdot)$ solves the martingale problem for operator $\mathcal{L}$ and is continuous. The $x(\cdot)$ in this paper will usually be the limit of a sequence in $D^r[0,\infty)$ (e.g., of the $x^\varepsilon(\cdot)$ introduced in Section 1), and <u>a-priori</u> we do not know that it is continuous, w.p.1. If it is not continuous at t_i or t or $t + s$, then we may not be able to show that (2.4) holds. But there are at most countably many points at which $x(\cdot) \in D^r[0,\infty)$ is not continuous w.p.1 : this set is $\mathcal{T}$. Henceforth we conveniently ignore $\mathcal{T}$.

<u>A truncation device</u>. In order to prove tightness, it is sometimes useful to work with a truncated process. Referring to (1.2), for each N, let $x^{\varepsilon,N}(\cdot)$ denote a process which equals $x^\varepsilon(\cdot)$ until first

exit from $S_N = \{x : |x| \leq N\}$. For each N, let there be $x^N(\cdot)$ such that $x^{\varepsilon,N}(\cdot) \Rightarrow x^N(\cdot)$, where $x^N(\cdot)$ solves the martingale problem for an operator $\mathscr{L}^N$ defined by

$$\mathscr{L}^N f(x) = f_x'(x) b^N(x) + \frac{1}{2} \sum_{i,j} a_{ij}^N(x) f_{x_i x_j}(x).$$

Let $b^N(x) = b(x)$ and $\sigma^N(x) = \sigma(x)$ for $x \in S_N$, and let the martingale problem for operator $\mathscr{L}$ have a unique solution. Then $x^\varepsilon(\cdot) \Rightarrow x(\cdot)$, the solution to the martingale problem for operator $\mathscr{L}$ [5,6].

Let $q_N(\cdot)$ denote a twice continuously differentiable function which equals 1 in S_N and zero on $R^r - S_{N+1}$. For the system (1.1), the N-truncation is defined by

$$X_{n+1}^{\varepsilon,N} = X_n^{\varepsilon,N} - \varepsilon Y_n (Y_n' X_n^{\varepsilon,N} - \psi_n) q_N(X_n^{\varepsilon,N}), \tag{2.5}$$

$$x^{\varepsilon,N}(t) = X_n^{\varepsilon,N} \text{ on } [n\varepsilon, n\varepsilon + \varepsilon),\ X_0^{\varepsilon,N} = X_0^\varepsilon.$$

Using these remarks,we will often simply proceed in the analysis as if $\{x^\varepsilon(\cdot)\}$ were bounded.

<u>Notation</u>. When no confusion arises, we write t/ε for the largest integer i such that $i \leq t/\varepsilon$ The symbol E_n denotes conditioning on $\{X_0^\varepsilon, Y_j, \psi_j, j < n\}$.

3. Weak Convergence of $\{x^\varepsilon(\cdot)\}$.

Theorem 1. Let $X_0^\varepsilon \Rightarrow x_0$. Assume (3.1) and either (3.2i) or (3.2ii) for some matrix R and vector B, as $n \to \infty$ and $N \to \infty$.

(3.1) $\{Y_n Y_n', Y_n \psi_n\}$ is uniformly integrable

(3.2i) $$\frac{1}{N} \sum_{j=n}^{n+N} Y_j Y_j' \xrightarrow{P} R, \quad \frac{1}{N} \sum_{j=n}^{n+N} Y_j \psi_j \xrightarrow{P} B$$

(3.2ii) $$\frac{1}{N} \sum_{j=n}^{n+N} E_n Y_j Y_j' \xrightarrow{P} R, \quad \frac{1}{N} \sum_{j=n}^{n+N} E_n Y_j \psi_j \xrightarrow{P} B$$

Then $x^\varepsilon(\cdot) \Rightarrow x(\cdot)$, which satisfies

(3.3) $$\dot{x} = -Rx + B, \; x(0) = x_0.$$

Remark. Under (3.1), (3.2ii) is implied by (3.2i). There are several approaches to obtaining (3.3), among them being the schemes in [8]. The method here has the advantage of being easy to generalize. The conditions are clearly related to those used in [6],[7],[8]. Reference [7] examines a recursive algorithm with 'state dependent' noise.

Proof. We work with the N-truncation and show $x^{\varepsilon,N}(\cdot) \Rightarrow x^N(\cdot)$ where

(3.4) $$\dot{x}^N = (-Rx^N + B)q_N(x^N), \; x^N(0) = x_0.$$

As noted in section 2, this implies that $x^\varepsilon(\cdot) \Rightarrow x(\cdot)$ satisfying (3.3). We use (3.2ii): under (3.2.i) the proof is similar. Let n_ε satisfy $n_\varepsilon \to \infty$ and $\varepsilon n_\varepsilon \equiv \delta_\varepsilon \to 0$ as $\varepsilon \to 0$. Fix N. By the truncation and (3.1), $\{Y_n(Y_n' X_n^{\varepsilon,N} - \psi_n) q_N(X_n^{\varepsilon,N})\}$ is uniformly integrable, so $\{x^{\varepsilon,N}(\cdot)\}$ is tight in $D^r[0,\infty)$, and all (weak) limits lie in $C^r[0,\infty)$. Fix a convergent subsequence (also indexed by ε) with limit $x^N(\cdot)$. Fix $f \in \hat{C}_0^2$ and define $g(X,Y,\psi) = f_x'(X)\cdot[-Y(Y'X - \psi)]q_N(X)$. <u>Henceforth we suppress the</u> N <u>superscript on</u> $x^{\varepsilon,N}(t)$ <u>and</u> $X_j^{\varepsilon,N}$, <u>but retain it for the limit</u> $x^N(\cdot)$. Define the piecewise constant function $g^\varepsilon(\cdot)$ by

$$g^\varepsilon(t) = \frac{1}{n_\varepsilon} \sum_{j=\ell n_\varepsilon}^{\ell n_\varepsilon + n_\varepsilon - 1} E_{\ell n_\varepsilon} g(X_j^\varepsilon, Y_j, \psi_j) \quad \text{on} \quad [\ell\delta_\varepsilon, \ell\delta_\varepsilon + \delta_\varepsilon).$$

Fix ℓ, $t_i < t < t + s$, $i \le \ell$, and let $h(\cdot)$ be bounded and continuous. There are Δ_i^ε such that $|\Delta_i^\varepsilon| \to 0$ as $\varepsilon \to 0$ and

$$(3.5) \quad Eh(x^\varepsilon(t_i), i \le \ell)[f(x^\varepsilon(t+s)) - f(x^\varepsilon(t)) - \varepsilon \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon} g(X_j^\varepsilon, Y_j, \psi_j)] = \Delta_1^\varepsilon$$

Also, by the properties of conditional expectations, since $t_i < t$,

$$(3.6) \quad Eh(x^\varepsilon(t_i), i \le \ell)[f(x^\varepsilon(t+s)) - f(x^\varepsilon(t)) - \delta_\varepsilon \sum_{\ell=t/\delta_\varepsilon}^{(t+s)/\delta_\varepsilon} g^\varepsilon(\ell\delta_\varepsilon)] = \Delta_2^\varepsilon,$$

$$(3.7) \quad Eh(x^\varepsilon(t_i), i \le \ell)[f(x^\varepsilon(t+s)) - f(x^\varepsilon(t)) - \int_t^{t+s} g^\varepsilon(u)du] = \Delta_3^\varepsilon.$$

Below, it will be shown that

$$(3.8) \quad g^\varepsilon(v) \xrightarrow{P} f_x'(x^N(v)[-Rx^N(v) + B]q_N(x^N(v)), \text{ each } v, \text{ as } \varepsilon \to 0.$$

Assume (3.8) for the moment. By the weak convergence and Skorohod imbedding, $\{x^\varepsilon(t_i),\ i \le \ell,\ x^\varepsilon(t),\ x^\varepsilon(t+s)\} \to \{x^N(t_i),\ i \le \ell,\ x^N(t),\ x^N(t+s)\}$ w.p.1. Using this, (3.8) and the fact that $\sup_{\varepsilon, v \le t} E|g^\varepsilon(v)| < \infty$, and taking limits in (3.7) yields

$$\text{(3.9)}\quad Eh(x^N(t_i),\ i \le \ell)[f(x^N(t+s)) - f(x^N(t)) - \int_t^{t+s} du\, f_x'(x^N(u))[-Rx^N(u) + B]q_N(x^N(u))] = 0.$$

Eqn. (3.9) and the arbitrariness of $f(\cdot)$, $h(\cdot), \ell, t_i, t$ and $t+s$, imply that $x^N(\cdot)$ solves the martingale problem for the operator $\mathscr{L}^N$ defined by

$$\mathscr{L}^N f(x) = f_x'(x)[-Rx + B]q_N(x).$$

Thus (3.4) holds, and we need only prove (3.8).

Fix v and let $\ell_\varepsilon \delta_\varepsilon \to v$ as $\varepsilon \to 0$. Define $m_\varepsilon = \ell_\varepsilon n_\varepsilon$. By the weak convergence and Skorohod imbedding $x^\varepsilon(\cdot) \to x^N(\cdot)$, a continuous process. Thus $X_j^\varepsilon \to x^N(v)$ uniformly for $j \in [\ell_\varepsilon n_\varepsilon, \ell_\varepsilon n_\varepsilon + n_\varepsilon)$ as $\varepsilon \to 0$. Thus, by the continuity of $f_x(\cdot)$ and $q_N(\cdot)$ and (3.1),

$$\lim_{\varepsilon\to 0} g^\varepsilon(v) = \lim_{\varepsilon\to 0} \frac{1}{n_\varepsilon} \sum_{j=m_\varepsilon}^{m_\varepsilon+n_\varepsilon-1} E_{m_\varepsilon} f_x'(X_j^\varepsilon)[-Y_j(Y_j'X_j^\varepsilon - \psi_j)]q_N(X_j^\varepsilon)$$

$$\text{(3.10)}\qquad = \lim_{\varepsilon\to 0} \frac{1}{n_\varepsilon} \sum_{j=m_\varepsilon}^{m_\varepsilon+n_\varepsilon-1} E_{m_\varepsilon} f_x'(X_{m_\varepsilon}^\varepsilon)[-Y_j(Y_j'X_{m_\varepsilon}^\varepsilon - \psi_j)]q_N(X_{m_\varepsilon}^\varepsilon)$$

$$= q_N(x^N(v))f_x'(x^N(v)) \lim_{\varepsilon\to 0} \frac{1}{n_\varepsilon} \sum_{j=m_\varepsilon}^{m_\varepsilon+n_\varepsilon-1} E_{m_\varepsilon}[-Y_j(Y_j'X_{m_\varepsilon}^\varepsilon - \psi_j)],$$

where all limits are in probability. Finally, by (3.2ii) and the weak convergence, the last line of (3.10) yields (3.8).

Q.E.D.

4. Limits as $\varepsilon \to 0$ and $\varepsilon n \to \infty$.

Since the system (1.1) is usually in operation for a long time, the behavior as $\varepsilon \to 0$ and $\varepsilon n \to \infty$ simultaneously is of considerable interest. Define $\tilde{x}^\varepsilon(\cdot) = x^\varepsilon(t_\varepsilon + \cdot)$, where $t_\varepsilon \to \infty$ as $\varepsilon \to 0$. Weak convergence alone does not imply that $\tilde{x}^\varepsilon(\cdot) \Rightarrow$ stationary solution to (3.3). For example, define $y^\varepsilon(t) = \max[0, t - 1/\varepsilon]$. Then $y^\varepsilon(\cdot) \Rightarrow$ zero function, but $\lim_t y^\varepsilon(t) \to \infty$ for each ε. However, under slightly stronger conditions than used in the last section, we do get the desired limit as $\varepsilon \to 0$, $t_\varepsilon \to \infty$. First, we prove Theorem 2, and then a criterion for the tightness will be given. If $R > 0$, define $\theta = R^{-1}B$.

Theorem 2. Let $\{X_j^\varepsilon, \varepsilon > 0, j < \infty\}$ be tight in R^r, and assume (3.1) and either (3.2i or ii). Then $\{\tilde{x}^\varepsilon(\cdot)\}$ is tight and all weak limits satisfy (3.3). If $R > 0$, then the weak limit is the constant function with value θ, the stationary solution.

Proof outline. The first assertion follows from the proof of Theorem 1. We need only characterize the limit process when $R > 0$. To do this, we exploit the stability of (3.3). For any $T < \infty$, take a weakly convergent subsequence of the pair $\{\tilde{x}^\varepsilon(\cdot), \tilde{x}^\varepsilon(\cdot - T)\}$, with limit $(x(\cdot), x_T(\cdot))$. We have $x(0) = x_T(T)$. The set of possible $\{x_T(0)\}$ is tight (over all T and convergent subsequences), since $\{X_j^\varepsilon, \varepsilon > 0, j < \infty\}$ is. Thus, the conclusion is implied by the arbitrariness of T and the representation

$$x_T(T) = x(0) = (\exp - RT)x_T(0) + \int_0^T (\exp - R(T - t))Bdt =$$

$$(\exp - RT)x_T(0) + \theta - \int_T^\infty (\exp - Rt)Bdt.$$

The next proof illustrates the use of the 'perturbed Liapunov function' method [6],[9],[10]. It exploits the stability of (3.3) to obtain tightness of the iterates and of the sequence of normalized errors.

Theorem 3. Let $R > 0$ and let $\{Y_j,\psi_j\}$ be bounded. Suppose that

$$\sum_{j=n}^{\infty} |E_n Y_j Y_j' - R|, \quad \sum_{j=n}^{\infty} |E_n Y_j \psi_j - B|$$

is bounded, uniformly in n and let $\{X_0^\varepsilon\}$ be tight. Then there are $N_\varepsilon < \infty$ such that

(4.1) $\{X_j^\varepsilon,\ \varepsilon$ small, $j < \infty\}$ is tight

(4.2) $\{(X_j^\varepsilon - \theta)/\sqrt{\varepsilon},\ \varepsilon$ small, $j \geq N_\varepsilon\}$ is tight.

(If $(X_j^\varepsilon - \theta) = O(\sqrt{\varepsilon})$, we can set $N_\varepsilon = 0$).

Proof. Recall that $\theta = R^{-1}B$ and define $\delta_n^\varepsilon = X_n^\varepsilon - \theta$. Then

$$(4.3)\quad \delta_{n+1}^\varepsilon = \delta_n^\varepsilon - \varepsilon(Y_n Y_n' - R)\delta_n^\varepsilon - \varepsilon R \delta_n^\varepsilon + \varepsilon(Y_n\psi_n - Y_n Y_n'\theta).$$

Henceforth, suppress the ε superscript on δ_n^ε and suppose without loss of generality that $\{X_0^\varepsilon\}$ is uniformly bounded. Define $V(\delta) = \delta'\delta$ and define the function $W^\varepsilon(n)$ by

$$W^\varepsilon(n) = 2\varepsilon\delta_n' \sum_{j=n}^{\infty} E_n(Y_j Y_j' - R)\delta_n - 2\varepsilon\delta_n' \sum_{j=n}^{\infty} E_n(Y_j\psi_j - Y_j Y_j'\theta).$$

The K below are real numbers which do not depend on ε and whose values might change from usage to usage. Define the perturbed Liapunov function $V^\varepsilon(n) = V(\delta_n) - W^\varepsilon(n)$.

We have

$$
\begin{aligned}
&|W^\varepsilon(n)| \le K\varepsilon[1 + V(\delta_n)] \\
&|W^\varepsilon(n)| \le K\varepsilon[1 + |V^\varepsilon(n)|].
\end{aligned}
\tag{4.4}
$$

Also

$$
\begin{aligned}
E_n V(\delta_{n+1}) - V(\delta_n) &= -2\varepsilon\delta_n' R\delta_n \\
&\quad -2\varepsilon\delta_n' E_n(Y_n Y_n' - R)\delta_n + 2\varepsilon\delta_n' E_n(Y_n\psi_n - Y_n Y_n'\theta) \\
&\quad + O(\varepsilon^2)(1 + |\delta_n|^2),
\end{aligned}
$$

$$
\begin{aligned}
E_n W^\varepsilon(n+1) - W^\varepsilon(n) &= -2\varepsilon\delta_n' E_n(Y_n Y_n' - R)\delta_n \\
&\quad + 2\varepsilon\delta_n' E_n(Y_n\psi_n - Y_n Y_n'\theta) + O(\varepsilon^2)[|\delta_n| + |\delta_n|^2].
\end{aligned}
$$

Thus there is a $\lambda > 0$ such that for small $\varepsilon > 0$,

$$
\begin{aligned}
E_n^\varepsilon V^\varepsilon(n+1) - V^\varepsilon(n) &= -2\varepsilon\delta_n' R\delta_n + O(\varepsilon^2)[1 + |\delta_n|^2] \\
&\le -\varepsilon\lambda V^\varepsilon(n) + O(\varepsilon^2),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
&EV^\varepsilon(n) \le (1 - \varepsilon\lambda)^n EV^\varepsilon(0) + O(\varepsilon) \\
&EV(\delta_n) \le K(1 - \varepsilon\lambda)^n EV(\delta_0) + O(\varepsilon).
\end{aligned}
\tag{4.5}
$$

The theorem follows from (4.5). Q.E.D.

5. Local Behavior Near θ.

Under the conditions of Theorem 3, $\overline{\lim_n} E|X_n^\varepsilon - \theta|^2 = O(\varepsilon)$. In fact, we can do much better. Define $U_n^\varepsilon = (X_n - \theta)/\sqrt{\varepsilon}$. Using any N_ε satisfying the needs of Theorem 3, define $u^\varepsilon(\cdot)$ by $u^\varepsilon(0) = U_{N_\varepsilon}^\varepsilon$ and $u^\varepsilon(t) = U_{N_\varepsilon+j}^\varepsilon$ on $[j\varepsilon, j\varepsilon + \varepsilon)$. Let $t_\varepsilon \to \infty$ as $\varepsilon \to 0$ and set $\tilde{u}^\varepsilon(t) = u^\varepsilon(t + t_\varepsilon)$. Theorem 4 shows that $\{u^\varepsilon(\cdot)\}$ converges weakly to a certain Gaussian diffusion. The properties of this diffusion are quite helpful for our understanding of the effects of the noise and stability properties on the algorithm.

Theorem 4. Define $\xi_n = (Y_n\psi_n - Y_nY_n'\theta)$, and let

$$W_n^\varepsilon(t) = \sum_{j=n}^{n+t/\varepsilon} \sqrt{\varepsilon}\, \xi_j \tag{5.1}$$

converge weakly to a (possibly non-standard) Wiener process $w(\cdot)$ with covariance Σt, as $\varepsilon \to 0$ and $n \to \infty$. Assume (3.2i or ii) and the conditions of Theorem 3. Then $U_{N_\varepsilon}^\varepsilon$ is tight and each subsequence of $\{u^\varepsilon(\cdot)\}$ (resp., $\{\tilde{u}^\varepsilon(\cdot)\}$) contains a further subsequence which converges to a solution to (5.2) (resp., to the stationary solution to (5.2)). If $\{U_0^\varepsilon\}$ is tight, then we can use $N_\varepsilon = 0$ and the assertion concerning convergence of $\{u^\varepsilon(\cdot)\}$ remains valid if the conditions of Theorem 3 are replaced by (3.1).

$$du = -Rudt + Bdt + dw \tag{5.2}$$

Remark. Assuming the weak convergence (5.1) is much more convenient than stating conditions which guarantee it, since such convergence is the subject of a large literature. See, e.g., [1].

Proof. The proof is basically a modification of that of Theorem 1. We have

$$U_{n+1}^{\varepsilon} = U_n^{\varepsilon} - \varepsilon Y_n Y_n' U_n^{\varepsilon} + \sqrt{\varepsilon}\, \xi_n. \tag{5.3}$$

The tightness of $\{U_{N_\varepsilon}^{\varepsilon}\}$ follows from Theorem 3. For notational simplicity, set $N_\varepsilon = 0$. To prove the theorem properly, we should use the N-truncation $u^{\varepsilon,N}(\cdot)$ of $u^{\varepsilon}(\cdot)$ which is defined by $U_{n+1}^{\varepsilon,N} = U_n^{\varepsilon,N} - \varepsilon Y_n Y_n' U_n^{\varepsilon,N} q_N(U_n^{\varepsilon,N}) + \sqrt{\varepsilon}\xi_n$, and then show that $u^{\varepsilon,N}(\cdot) \Rightarrow u^N(\cdot)$, where $u^N(\cdot)$ satisfies $du^N = -Ru^N q_N(u^N)dt + Bdt + dw$, and then let $N \to \infty$ to obtain (5.2). But, in order to save notation and a few details, we ignore the truncation, and simply suppose that the $\{U_n^{\varepsilon}\}$ are bounded†. By this boundedness and the uniform integrability of $\{Y_n Y_n', Y_n \psi_n\}$ and the convergence $W_n^{\varepsilon}(\cdot) \Rightarrow w(\cdot)$, we have that $\{\varepsilon \sum_0^{t/\varepsilon} Y_j Y_j' U_j^{\varepsilon}\}$ and $\{W_n^{\varepsilon}(\cdot)\}$ are tight and that all limits have continuous paths. Thus $\{u^{\varepsilon}(\cdot)\}$ is tight, and all limits have continuous paths.

† Actually, the $\{U_j^{\varepsilon,N}\}$ are not bounded, but for each $T < \infty$ we have $\lim_{K\to\infty} \overline{\lim_{\varepsilon}} P\{\sup_{j\varepsilon \le T} |U_j^{\varepsilon,N}| \ge K\} = 0$. This and the cited uniform integrability and convergence of $W_n^{\varepsilon}(\cdot)$ to $w(\cdot)$ are enough to get tightness of $\{u^{\varepsilon,N}(\cdot)\}$, with all limits being in $C^r[0,\infty)$.

Define the functions $g^{\varepsilon}(t)$ and $G^{\varepsilon}(t)$ by

$$g^{\varepsilon}(t) = \frac{1}{n_{\varepsilon}} \sum_{\ell n_{\varepsilon}}^{\ell n_{\varepsilon}+n_{\varepsilon}-1} E_{\ell n_{\varepsilon}}(-Y_jY_j'U_j^{\varepsilon} + Y_j\psi_j), \text{ on } [\ell\delta_{\varepsilon}\ \ell\delta_{\varepsilon} + \delta_{\varepsilon}),$$

$$G^{\varepsilon}(t) = \int_0^t g^{\varepsilon}(u)du.$$

First, we work with $\{u^{\varepsilon}(\cdot)\}$.

We have

$$(5.4)\quad u^{\varepsilon}(t) - u^{\varepsilon}(0) = G^{\varepsilon}(t) + W_0^{\varepsilon}(t) + [\varepsilon \sum_{j=0}^{t/\varepsilon-1}(-Y_jY_j'U_j^{\varepsilon} + Y_j\psi_j) - G^{\varepsilon}(t)].$$

The process in the bracket in (5.4) is tight and converges weakly to the zero process, as $t \to \infty$. This can be shown by a slight modification of the method of Theorem 1. The sequence $\{u^{\varepsilon}(\cdot),W_0^{\varepsilon}(\cdot)\}$ is tight. Extract a weakly convergent subsequence with limit $(u(\cdot),w(\cdot))$, and indexed also by ε. By the method of Theorem 1, $G^{\varepsilon}(t) \xrightarrow{P} \int_0^t(-Ru(s)+B)ds$ for each t.

The limits of the different convergent subsequences differ only in the initial condition $u(0)$. The argument for $\{\tilde{u}^{\varepsilon}(\cdot)\}$ is the same as that for $\{u^{\varepsilon}(.)\}$, except for the stationarity part. But this part follows from an argument like that used in Theorem 2. The last assertion follows from the previous ones, since if $\{U_0^{\varepsilon}\}$ is tight, we do not need to prove the existence of $N_{\varepsilon} < \infty$, and so the boundedness of $\{Y_nY_n', Y_n\psi_n\}$ can be replaced by (3.1) and a truncation argument, when working with $\{u^{\varepsilon}(\cdot)\}$. Q.E.D.

6. Tracking Parameter Variations.

If the statistics of Y_j, ψ_j change with time and their "rate of change' is commensurate with ε, then Theorem 1 can be extended. Let there be continuous $R(\cdot)$ and $B(\cdot)$ such that for each t, (3.2) holds if $n\varepsilon \to t$ and $R(t)$ and $B(t)$, replace R and B, resp. Then, under (3.1) Theorem 1 continues to hold.

7. A Projection or Truncation Algorithm.

In practical problems, the iterates $\{X_n^\varepsilon\}$ are usually prevented from becoming too large by using a projection or truncation, and we now treat a simple case. Define the box $H = \{x : |x_i| \le k, i \le r\}$. Let $\pi_H(x)$ denote the closest point in H to x, and let $\pi(x,h)$ denote the projection onto H defined by

$$\pi(x,h) = \lim_{\Delta \to 0}[\pi_H(x + \Delta h) - x]/\Delta.$$

Thus $\pi(x,h) = h$ if x is interior to H, or if h points 'inside' when $x \in \partial H$, the boundary of H. We treat the algorithm

$$X_{n+1}^\varepsilon = \Pi_H[X_n^\varepsilon - \varepsilon Y_n(Y_n' X_n^\varepsilon - \psi_n)], \; X_0^\varepsilon = x_0 \in H. \tag{7.1}$$

Define $\delta x = x - \theta = x - R^{-1}B$, and write $F(x) = E(Y_n' x - \psi_n)^2$. If $\{Y_n, \psi_n\}$ is stationary and $R > 0$, $F(x)$ is strictly convex. The constraints that define the box H can be written as $q_i(x) \le 0$, $i = 1, \ldots, 2r$, where the $q_i(x)$ are of the form $x_i - k$ and $-k - x_i$. A point $x \in H$ is said to be a Kuhn-Tucker point iff there are $\lambda_i \geq 0$ such that

$$F_x(x) + \sum_{i \in A(x)} \lambda_i q_{i,x}(x) = 0,$$

where $A(x)$ denotes the set of constraints which are active at x. In the present case, x being a Kuhn-Tucker point is necessary and sufficient for its minimizing $F(\cdot)$ on H.

Theorem 5. Assume (3.1) and either (3.2i) or (3.2ii). Then $x^\varepsilon(\cdot) \Rightarrow x(\cdot)$, where

$$\dot{x} = \pi(x, -Rx + B),\ x_0 = x(0), \tag{7.2}$$

or, equivalently,

$$\delta\dot{x} = \pi(x, -R\delta x). \tag{7.3}$$

Let $R > 0$ and let $\{Y_n, \psi_n\}$ be stationary, and define $\tilde{x}^\varepsilon(\cdot)$ as in Theorem 2, then $\tilde{x}^\varepsilon(\cdot) \Rightarrow$ stationary solution to (7.3), which is a constant solution and a Kuhn-Tucker point for the problem of minimizing $F(x)$ subject to $x \in H$.

Proof. Rewrite (7.1) in the form

$$X^\varepsilon_{n+1} = X^\varepsilon_n - \varepsilon Y_n(Y'_n X^\varepsilon_n - \psi_n) + \varepsilon z^\varepsilon_n,$$

where

$$\varepsilon z^\varepsilon_n = \Pi_H[X^\varepsilon_n - \varepsilon Y_n(Y'_n X^\varepsilon_n - \psi_n)] - [X^\varepsilon_n - \varepsilon Y_n(Y'_n X^\varepsilon_n - \psi_n)].$$

Thus $\{z_n^\varepsilon, \varepsilon > 0, n < \infty\}$ is uniformly integrable. Using the notation of Theorem 1, define the functions $z^\varepsilon(\cdot)$ and $Z^\varepsilon(\cdot)$ by

$$z^\varepsilon(t) = \frac{1}{n_\varepsilon} \sum_{\ell n_\varepsilon}^{\ell n_\varepsilon + n_\varepsilon - 1} E_{\ell n_\varepsilon} f_x'(X_j^\varepsilon) z_n^\varepsilon \quad \text{on} \quad [\ell\delta_\varepsilon, \ell\delta_\varepsilon + \delta_\varepsilon),$$

$$Z^\varepsilon(t) = \int_0^t z^\varepsilon(u)du.$$

Define $g^\varepsilon(\cdot)$ and $G^\varepsilon(\cdot)$ as in Theorem 1. Since $\{X_n^\varepsilon\}$ is bounded by the projection, no N-truncation is needed. By (3.1), $\{x^\varepsilon(\cdot), Z^\varepsilon(\cdot)\}$ is tight and all limits are continuous. Also, the limits of $\{Z^\varepsilon(\cdot)\}$ are all absolutely continuous because of the uniform integrability of $\{z_n^\varepsilon\}$. Choose a weakly convergent subsequence indexed by ε and with limit $(x(\cdot), Z(\cdot))$. Define $z(\cdot)$ by $Z(t) = \int_0^t z(u)du$. By the method of Theorem 1

$$x(t) = x(0) + \int_0^t (-Rx(u) + B)du + \int_0^t z(u)du.$$

Note that $z_n^\varepsilon = 0$ if X_{n+1}^ε is interior to H. Also, if $X_{n+1}^\varepsilon \in \partial H$, then z_n^ε is a non-negative linear combination of the inward normals to the 'active surfaces' at $x = X_{n+1}^\varepsilon$. This implies the following *for almost all* t. If $x(t)$ is interior to H, then $z(t) = 0$. If $x(t) \in \partial H$, then $z(t)$ is a non-negative linear combination of the inward normals to the 'active surfaces' at $x = x(t)$. From this and the geometry of H, we conclude that $x(\cdot)$ satisfies (7.2) or (7.3).

We now discuss the stability of (7.3). Let $R > 0$ and define $V(x) = \delta x' R \delta x$. Then $\dot V(\delta x) = 2\delta x' R \pi(x, -R\delta x)$. If $v(\cdot)$ is continuous, then the function $K(v(x)) = v(x)'\pi(x, v(x))$

is non-positive and upper semicontinuous in v. It is zero only at those x where $\pi(x,v(x)) = 0$. Thus $\delta x(t)$ converges to the point δx where $\pi(x,-R\delta x) = 0$. This δx is a Kuhn-Tucker point since either (a): x is interior to H, in which case $\delta x = 0$ and $R\delta x = 0$, or (b): $x \in \partial H$, in which case $R\delta x$ must be a non-negative linear combination of the inward normals of the constraints which are active at x. From this point on, the proof of the convergence of $\{\tilde{x}^{\varepsilon}(\cdot)\}$ to the stationary solution (the Kuhn-Tucker point) is essentially the same as the convergence proof for $\{\tilde{x}^{\varepsilon}(\cdot)\}$ in Theorem 2 and is omitted. Q.E.D.

REFERENCES

1. P. Billingsley, Convergence of Probability Measures, 1968, Wiley, New York.

2. T.G. Kurtz, Approximation of Population Processes, vol. 36, CDMS/NSF series in Applied Math. SIAM, 1981.

3. A.V. Skorohod, "Limit theorems for stochastic processes", Theory of Prob. and its Applic., 1, 1956, 261-290.

4. D.W. Stroock, S.R.S. Varadhan, Multidimensional Diffusion Processes, 1979, Springer, Berlin.

5. H.J. Kushner, "A martingale method for the convergence of a sequence of processes to a jump-diffusion process", Z. Wahr., 53 (1980), 207-219.

6. H.J. Kushner, Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory, to be published by M.I.T. Press, Cambridge, U.S.A.

7. H.J. Kushner, A. Shwartz, "An invariant measure approach to the convergence of stochastic approximations with state dependent noise", to appear SIAM J. on Control and Optimization.

8. H.J. Kushner, D.S. Clark, Stochastic Approximation Methods for Constrained and Unconstrained Systems, 1978, Springer-Verlag, Berlin.

9. B.C. Blankenship, G.C. Papanicolaou, "Stability and control of stochastic systems with wide band noise disturbances", SIAM J. Appl. Math. 34, 1978, 437-476.

10. H.J. Kushner, Hai Huang, "Averaging methods for the asymptotic analysis of learning and adaptive systems with small adjustment rate", SIAM J. on Control and Optim., 19 (1981), 635-650.

11. W. Zangwill, Nonlinear Programming, A Unified Approach, 1969, Prentice Hall, Englewood Cliffs, N.J.

DATE
FILMED