

AD-A033 789

BROWN UNIV PROVIDENCE R I DIV OF APPLIED MATHEMATICS F/G 12/2
CONVERGENCE OF RECURSIVE ADAPTIVE AND IDENTIFICATION PROCEDURES--ETC(U)
1976 H J KUSHNER N00014-76-C-0279

UNCLASSIFIED

AFOSR-TR-76-1410

NL

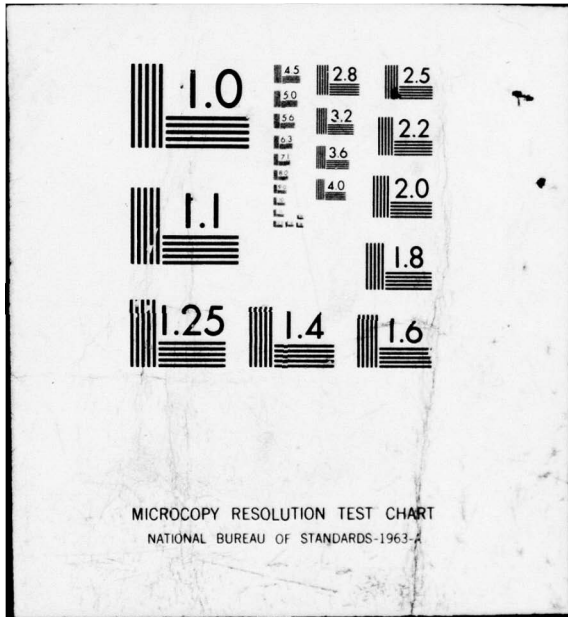
| OF |

AD
A033789



END

DATE
FILMED
2-77



3

CONVERGENCE OF RECURSIVE ADAPTIVE AND IDENTIFICATION PROCEDURES VIA WEAK CONVERGENCE THEORY

B.S.

Harold J. Kushner⁺

See 1473

Abstract

Results and concepts in the theory of weak convergence of a sequence of probability measures are applied to convergence problems for a variety of recursive adaptive (stochastic approximation like) methods. Similar techniques have had wide applicability in areas of operations research and in some other areas in stochastic control. It is quite likely that they will play a much more important role in control theory than they do at present, since they allow relatively simple and natural proofs for many types of convergence and approximation problems. Part of the aim of the paper is tutorial: to introduce the ideas, and to show how they might be applied. Also, many of the results are new, and they can all be generalized in many directions.

ADA033789

Approved for public release;
distribution unlimited.

DDC
RECEIVED
DEC 28 1976
D

⁺Division of Applied Mathematics and Engineering, Brown University, Providence, R.I. This work was partially supported by the Air Force Office of Scientific Research under AF-AFOSR-76-3063, by the National Science Foundation under Eng-73-03846-A01, and the Office of Naval Research under NONR N000 14-76-C-0279.

3

CONVERGENCE OF PROBABILISTIC APPROXIMATION
PROCEDURES FOR REAL CONVEX OPTIMIZATION

Arnold J. Gashner

Abstract

Results and concepts in the theory of weak convergence of a sequence of probability measures are applied to convergence problems for a variety of recursive stochastic approximation (stochastic approximation like) methods. Similar techniques have had wide application in areas of operations research and in some other areas in stochastic control. It is quite likely that they will play a much more important role in control theory than they do at present, since they allow relatively simple and natural proofs for many types of convergence and approximation problems. Part of the aim of this paper is to illustrate the technique in this way and to show how they might be applied. Also, many of the results are new, and they can all be generalized in many directions.

DDC
RECEIVED
DEC 28 1972

Approved for Release
Distribution Unlimited

Division of Applied Mathematics and Engineering, Brown University, Providence, R.I. This work was partially supported by the Air Force Office of Scientific Research, Durham, N.C.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

88-38038

1. Introduction

The aims of this paper are two-fold. The first aim is tutorial. The technique of and the results in the theory of weak convergence of a sequence of probability measures have found many useful applications in many areas of operations research and statistics [1], [2]. Their role in control theory has been relatively limited, being confined mainly to the work in [3], [4] which deal with control problems on diffusion models. Yet, its intrinsic power as well as the nature of the past successes, suggests that its role in control theory should be deeper than it is at present. The techniques are particularly valuable when convergence or approximation ideas are being dealt with.

In order to illustrate the possibilities, the ideas of weak convergence theory will be applied (the second goal of the paper) to some convergence problems for an interesting class of adaptive processes. These processes have the interesting stochastic approximation (SA) like framework used by Ljung and others [5], [6], and a number of practical applications. The application to the convergence problem will illustrate some of the main ideas of weak convergence theory. Some of Ljung's results will be rederived.

Sometimes our conditions are weaker, and sometimes stronger. Our proofs are generally much simpler. They appear to be readily generalizable to more abstract cases, and conditions on the noise and coefficient sequences are weaker. The ideas used

here allow simpler proofs, and focus on somewhat different types of conditions. They are essentially "invariant" with respect to "perturbations". Other advantages will be discussed in the sequel. For example, there are extensions to the case where state space constraints must be included. The results here do not replace those of Ljung. Both methods (which are not unrelated) are quite interesting, and various combinations of them may well prove more fruitful than either one alone, in allowing us to handle broader classes of SA like procedures - which include more realistic noise processes, etc.

The classical techniques are too cumbersome and too over-dependent on special properties - such as square summability of certain coefficient sequences, and orthogonality properties of the noise. "On line" methods of identification, for example, usually require rather weak assumptions on the noise sequences. In any case, more powerful methods for the handling of such recursive algorithms have been long needed.

In Section 2, some of the ideas of weak convergence theory are introduced. Section 3 elaborates certain points and criteria of Section 2, Section 4 develops the main application, and certain extensions are discussed in Section 5.

To motivate our point of view, suppose that $\{X^n(\cdot), T_1 \leq t \leq T_2\}$ (possibly $T_2 = \infty$ and/or $T_1 = -\infty$) is a sequence of random processes whose paths are in a path or function space \mathcal{A} , w.p.1, for each n . It turns out that if we view each $X^n(\cdot)$ as an abstract valued random variable (with

values in \mathcal{X}), and study the sequence of measures induced on \mathcal{X} by $\{x^n(\cdot)\}$, then very useful results can often be obtained on various limiting ($n \rightarrow \infty$) properties of the sequence. For this reason, it is useful to study sequences of probabilities on suitable abstract spaces, even if the applications are concerned.

ACCESSION for	
RTIS	Write Section <input checked="" type="checkbox"/>
BUG	Buff Section <input type="checkbox"/>
UNANNOUNCED	<input type="checkbox"/>
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	AVAIL. CODE/SP. CODE
A	

DDC
 RECEIVED
 DEC 28 1976
 D

2. Weak Convergence of Measures

The main reference is Billingsley [7]. See also Gikhman and Skorokhod [8], Chapter 9, or Chapter 2 in Kushner [4], for a brief summary of the basic ideas. Weak convergence is a generalization to abstract valued random variables of convergence in distribution. The statements below (unless otherwise specified) are in [7], Chapter 1. Let \mathcal{X} denote a complete separable metric space. Suppose for the moment that the processes⁺ are of interest over a finite interval $[T_1, T_2]$. Then \mathcal{X} is usually taken to be $C[T_1, T_2]$ or $D[T_1, T_2]$ (or C^m, D^m , their m -fold products), where C is the space of real-valued continuous functions with the sup norm, and D is the space of real-valued functions which are right continuous, continuous at T_2 , and have left hand limits on $(T_1, T_2]$. The space D is often much more convenient to work with than the space C , but, for simplicity only, C will be used here. (See [7], or [4], Chapter 2, for a discussion of the topology which is usually used on D .) If T_1 or T_2 are infinite, then the usual extension of the topology on C is used (convergence is then equivalent to uniform convergence on finite intervals).

Let $\{X^n\}$ denote a sequence of random variables with values in \mathcal{X} , let $\{P^n\}$ denote the corresponding induced measures on the (Borel) sets of \mathcal{X} , and let $C(\mathcal{X})$ (resp., $C_P(\mathcal{X})$, where P is a measure on \mathcal{X}) denote the set of real-valued continuous, bounded, functions on \mathcal{X} (resp., real-valued, bounded, measurable and continuous almost everywhere on \mathcal{X} , with respect to P). The

⁺The processes of concern are to be real or vector valued. \mathcal{X} is the space in which the paths lie - not where the values lie. A process $X(\cdot)$ is considered to be an \mathcal{X} valued random variable, where convenient.

sequence $\{P^n\}$ is said to converge weakly to P (written $P^n \Rightarrow P$) if

$$(2.1) \quad \int f(y)P^n(dy) \rightarrow \int f(y)P(dy)$$

for all $f(\cdot) \in C(\mathcal{X})$. If (2.1) holds for all such $f(\cdot)$, it also holds for all $f(\cdot) \in C_p(\mathcal{X})$. This is an important generalization, since many of the $f(\cdot)$ of common interest in control theory are not continuous everywhere⁺ (see examples in [4]). Clearly, if \mathcal{X} is an Euclidean space, then weak convergence is equivalent to convergence in distribution. Let $\mathcal{X} = C^m[T_1, T_2]$, and let P^n be a measure on \mathcal{X} induced by a process $X^n(\cdot)$ or random variable X^n . If P is a measure on \mathcal{X} , there is a separable R^m valued process $X(\cdot)$ on $[T_1, T_2]$ with continuous paths w.p.1, which induces P on \mathcal{X} . If $P^n \Rightarrow P$, we abuse terminology and say that $X^n(\cdot) \rightarrow X(\cdot)$ (or $X^n \rightarrow X$) weakly, or in distribution.

The sequence $\{P^n\}$ (or $\{X^n\}$) is said to be tight, if for each $\epsilon > 0$, there is a compact set $K_\epsilon \in \mathcal{X}$ such that

$$(2.2) \quad P^n\{K_\epsilon\} = P\{X^n(\cdot) \in K_\epsilon\} \geq 1 - \epsilon, \text{ all } n.$$

If $\{P^n\}$ is tight, then for each subsequence, there is a further subsequence (denoted by $\{P^{n'}\}$) and a measure P such that $P^{n'} \Rightarrow P$. Indeed, tightness is necessary and sufficient for $\{P^n\}$

⁺E.g., $f(\cdot)$ that relate to exit times of a process from a set.

to be relatively compact. A product sequence $\{P_1^n \times P_2^n\}$ (corresponding to, say, a sequence of pairs $\{X_1^n, X_2^n\}$) is tight if each component $\{P_i^n\}$ is tight.

Following the forementioned abuse of terminology, if $\{P^n\}$ is tight, we may say that $\{X^n\}$ is tight, then that there is a weakly convergent subsequence of $\{X^n\}$ with limit X (where, if $X^n(\cdot)$ is a process with paths in a C space, $X(\cdot) = X$ will be also).

In practice, the $\{X^n(\cdot)\}$ can arise in many ways. It may be a sequence of approximations to a process or optimal process $X(\cdot)$, which is obtained by (say) some computational procedure, and it may be desired to that $X^n(\cdot)$ converges to $X(\cdot)$ in some sense. In many examples, a problem or process may be parametrized by a scaling factor (as is often the case in applications to Queueing[1]) or other parameter α . A limit process ($\alpha = 0$ or $\alpha = \infty$) may be easy to study, and it may be desired to show that $X^\alpha \rightarrow X^0$ (or X^∞) in a suitable sense. In this paper, the process $X^n(\cdot)$ arise in a somewhat different way. See Section 4 on.

The method of attack is often as follows. First tightness of $\{X^n\}$ is proved by using one of the many available criteria. Then an arbitrary convergent subsequence is selected (of which there is at least one, by tightness). Using properties of the sequences (as done in Section 4), the limiting process $X(\cdot)$ is characterized. Then, we try to show that the character of limit does not depend on the subsequence. Finally, via (2.1) or similar results, we have convergence of various functionals of the

sequence to a functional of the limit - in distribution or expectation. One of the main advantages of the technique is that, once we know there is tightness, we know that we can extract and treat convergent subsequences. It is not necessary to prove that they exist - as will be seen below. This is a great advantage.

One of the most useful tools in applications of weak convergence theory is known as Skorokhod imbedding (see Skorokhod [9], Theorem 3.1.1 or [4], Theorem 2.2). The theorem is the following. Let P^n and P be induced (on \mathcal{X}) by the \mathcal{X} valued random variables X^n and X , resp., and let $P^n \Rightarrow P$. Then there is some probability space $(\tilde{\Omega}, \tilde{\mathcal{G}}, \tilde{P})$ with \mathcal{X} valued random variables $\{\tilde{X}^n\}$ and \tilde{X} defined on it such that, for each Borel set A in \mathcal{X} ,

$$(2.3) \quad \begin{aligned} P\{X^n \in A\} &= \tilde{P}\{\tilde{X}^n \in A\} \\ P\{X \in A\} &= \tilde{P}\{\tilde{X} \in A\} \end{aligned}$$

and $\tilde{X}^n \rightarrow \tilde{X}$ w.p.1 in the topology of \mathcal{X} .

Let $\mathcal{X} = C^m(-\infty, \infty)$, and let X^n and X be \mathcal{X} valued random variables, w.p.1. Thus, the corresponding processes $X^n(\cdot), X(\cdot)$, are defined on $(-\infty, \infty)$, are R^m valued, and have continuous paths w.p.1. Suppose that $P^n \Rightarrow P$. Let \tilde{X}^n and \tilde{X} (corresponding to R^m valued continuous processes $\tilde{X}^n(\cdot), \tilde{X}(\cdot)$) on $(\tilde{\Omega}, \tilde{P}, \tilde{\mathcal{G}})$ correspond to X^n and X , by the Skorokhod imbedding. Then $\tilde{X}^n \rightarrow \tilde{X}$ w.p.1 on \mathcal{X} , and this implies that $\sup_{|t| \leq T} |\tilde{X}^n(t) - \tilde{X}(t)| \rightarrow 0$ w.p.1, as $n \rightarrow \infty$, for each $T < \infty$. While the distributions of \tilde{X}^n (resp., of

\tilde{X}) are exactly those of X^n (resp., X), the probability spaces are different. Usually, we are concerned mainly with characterizing the limits P, X , and, since \tilde{X} is equivalent to X in the sense that it induces the same distributions on $C^r(-\infty, \infty)$ (or on whatever \mathcal{A} is), the properties of \tilde{X} often yield the desired properties of X . The imbedding allows us to use w.p.1 convergence in certain places. When the imbedding (and consequent change of probability space) is used here, it will be so stated - but the tilde ~ notation will not be used. The same symbols will be used for both the original and the ~ process.

Consider a simple example. Let X^n, X be real-valued with $P\{X^n = 1\} = 1 - P\{X^n = 0\} = \frac{1}{n}$, $P\{X = 0\} = 1$. Define $\tilde{\Omega} = [0, 1]$, $\tilde{\mathcal{G}}$ = Borel sets on $[0, 1]$, and \tilde{P} = Lebesgue measure on $[0, 1]$. Define \tilde{X}^n and \tilde{X} by: $\tilde{X}^n = 1$ on $[0, \frac{1}{n})$, and zero elsewhere, $\tilde{X} = 0$ on $[0, 1]$. Then (2.3) holds, and $\tilde{X}^n \rightarrow \tilde{X}$ w.p.1. The joint distributions of $\tilde{X}_1, \dots, \tilde{X}_n, \dots, \tilde{X}$ and of $X_1, \dots, X_n, \dots, X$ are not necessarily the same; but, if we are only concerned with the probabilistic properties of the limit X , we can just as well use the imbedded process. In fact, in many applications each X^n is defined on a different probability space anyway (but not in this paper), in which case the "joint distributions" of $\{X_1, \dots, X_n, \dots, X\}$ has no meaning anyway.

3. Criteria for Tightness When $\mathcal{X} = C^r(-\infty, \infty)$.

Let us specialize to the case $C^r[-T, T]$ (see Billingsley [7], Section 8, where $C[0, T]$ is treated, for details) One of the critical aspects of applying the theory is the establishment of reasonably readily verifiable criteria for tightness. Since we usually work with processes and their properties, and not with measures on \mathcal{X} , the criteria should be, if possible, in terms of available data on the processes.

Suppose that $\{x_n(\cdot)\}$ is a sequence of R^r valued continuous functions on $[-T, T]$. Then to any subsequence of $\{x_n(\cdot)\}$, there is a further subsequence which converges to an element of $C^r[-T, T]$, if and only if the sequence $\{x_n(\cdot)\}$ is bounded and equicontinuous (by the Arzela-Ascoli Theorem). Thus, as is well known, the compact sets of $C^r[-T, T]$ are sets of equibounded and equicontinuous functions. The criteria for (2.2) all imply that the paths of $x^n(\cdot)$ are bounded and equicontinuous, with a "high enough" probability, where the bounds and moduli of continuity are not dependent on n .

The sequence $\{x^n(\cdot)\}$ is tight if and only if, for each $\eta > 0$, there is an $N_\eta < \infty$ such that

$$(3.1) \quad P\{|x^n(\cdot)| \geq N_\eta\} \leq \eta, \text{ all } n$$

and, for each $\varepsilon > 0$, $\eta > 0$, there is a $\delta \in (0, 1)$ and an $n_0 < \infty$ such that

$$(3.2) \quad P\left\{ \sup_{\substack{|t-s| < \delta \\ -T \leq t, s \leq T}} |X^n(t) - X^n(s)| \geq \epsilon \right\} \leq \eta, \text{ for } n \geq n_0.$$

If, for each $\epsilon > 0$, $\eta > 0$, there is a $\delta \in (0,1)$ and n_0 such that

$$(3.3) \quad P\left\{ \sup_{s \leq t \leq s+\delta} |X^n(t) - X^n(s)| \geq \epsilon \right\} \leq \eta\delta, \text{ for } n \geq n_0 \text{ and } -T \leq s \leq s+\delta \leq T,$$

then (3.2) holds. Equation (3.3) is guaranteed if there is a real K and an $a > 0$, $b > 0$, such that

$$(3.4) \quad E|X^n(t) - X^n(s)|^a \leq K|t-s|^{1+b}, \text{ all } n.$$

For the case $C^r(-\infty, \infty)$, we only need to satisfy the criteria on each $[-T, T]$, where $N_\eta, K, a, b, n_0, \delta$ can also depend on T . Of course, in special applications, much work can be devoted to showing that (3.2) or (3.3) or (3.4) hold. Note that the criterion (3.4), for a fixed n , is simply Kolmogorov's criterion for the path continuity of a separable random process. Here the criteria - or (3.2) or (3.3) - must hold for all large n with the constants not depending on n . This is hardly surprising.

It is often easier to show tightness in a D space than in a C space, since the conditions are weaker for the former. Also, working with D , it is often possible to show that the limits are continuous anyway. To avoid more descriptions, we stick to C .

4. A General Adaptive Algorithm⁺

Let $\{\xi_n\}$ denote a sequence of $R^{r'}$ valued random variables, and $\{a_n\}$ a non-negative null sequence of real numbers satisfying

$$(4.1) \quad \sum_n a_n = \infty.$$

Let $Q(\cdot, \cdot)$ denote a continuous bounded function: $R^r \times R^{r'} \rightarrow R^r$, and consider the algorithm

$$(4.2) \quad X_{n+1} = X_n - a_n Q(X_n, \xi_n), \quad n \geq 0, \quad X_0 \text{ given.}$$

Many adaptive and identification procedures fit into the form (4.2). Unbounded or even discontinuous $Q(\cdot, \cdot)$ can also be treated. Specific additional assumptions will be introduced below. The aim here is solely to discuss the convergence properties of $\{X_n\}$, and to illustrate how the techniques can be used in the treatment of applications, but not to deal with more specific applications of (4.2). (A number of applications are discussed in [5], [6].) To do this, using the ideas of Sections 2 and 3, $\{X_n\}$ must be interpolated into a continuous time function. A natural interpolation is suggested by the form of (4.2); the " a_n " is a natural time interval for the interpolation.

⁺More classical SA procedures, with and without constraints, are treated in [10] by somewhat similar methods.

Define⁺

$$t_0 = 0, t_n = \sum_{i=0}^{n-1} a_i, m(t) = \max\{n: t_n \leq t\},$$

and define a process $X^0(\cdot)$ by $X^0(t) = X_0, t \leq 0, X^0(t) = X^0(t_n) - Q(X_n, \xi_n)(t-t_n)$ on $[t_n, t_{n+1}]$. Define $X^n(\cdot)$ ($n = 1, 2, \dots$) by: $X^n(t) = X^0(t_n+t)$ for $t \geq -t_n$, and equal to $X_0 = X^n(-t_n)$, for $t \leq -t_n$. Thus $X^0(\cdot)$ is a piecewise linear interpolation of $\{X_n\}$, with interpolation intervals $\{a_n\}$, and $X^n(\cdot)$ is a left shift of $X^0(\cdot)$ by t_n . The purpose of the sequence of left shifts is to (eventually) bring the "asymptotic part" of the $\{X_n\}$ into some finite time interval. Define the piecewise constant interpolations: $\bar{X}(t) = X_n$ on $[t_n, t_{n+1})$, $\bar{\xi}(t) = \xi_n$ on $[t_n, t_{n+1})$, with $\bar{X}(t) = X_0, \bar{\xi}(t) = \xi_0, t < 0$. Then

$$(4.3) \quad \begin{aligned} X^0(t) &= X_0 - I^0(t) \\ X^n(t) &= X^n(0) - I^n(t), \end{aligned}$$

where $I^n(\cdot)$ is defined by

⁺The $m(t)$ and t_n terms will be used frequently, since the theory requires us to work with the interpolated processes, but the properties of the sequences $\{\xi_n, X_n\}$ must be referred to constantly. So, we go back and forth between the interpolated process and the sequences. The $m(t)$ and t_n allow us to keep track of the times at which the values of one are the values of the other. They also are responsible for most of the notational difficulties in the paper.

$$\begin{aligned}
 I^n(t) &= \int_0^t Q(\bar{X}(t_n+s), \bar{\xi}(t_n+s)) ds, \quad t \geq -t_n \\
 &= I^n(-t_n), \quad t \leq -t_n.
 \end{aligned}$$

Before proceeding, let us introduce some additional assumptions.

(A1) $\sup_n P\{|X_n| \geq K\} \rightarrow 0$ as $K \rightarrow \infty$.

(A2) There are measurable real valued functions $g(\cdot), \theta(\cdot)$ such that $\theta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ and

$$|Q(x, y) - Q(x', y)| \leq g(y)\theta(|x-x'|).$$

Define $G^n(t) = \int_0^t g(\bar{\xi}(t_n+s)) ds$, $t \in (-\infty, \infty)$, and suppose that $\{G^n(\cdot)\}$ is tight on $C(-\infty, \infty)$.

(A3) There is a random variable ξ such that, for any bounded and continuous real valued function $f(\cdot)$:

$$E[f(\xi_{n+k}) | \xi_i, i \leq n] \rightarrow Ef(\xi), \text{ as } n, k \rightarrow \infty.$$

(A4) Define $A(\cdot)$ by $A(x) = EQ(x, \xi)$. (Note that $A(\cdot)$ is bounded and continuous.) Suppose that S , the set of zeroes of $A(\cdot)$, is bounded and connected, and that $\dot{x} = -A(x)$ is asymptotically stable (to S).

Without the connectedness property, the conclusions below (4.6) are to be replaced by: $X_n \xrightarrow{P}$ largest finite invariant set

of $\dot{x} = -A(x)$, and this invariant set replaces S in (4.7). (A1) can frequently be verified by a Liapunov function type of method, or $Q(\cdot, \cdot)$ may be constructed to have some type of intrinsic stability. The aim here is to show the convergence, assuming that the mass does not wander off to infinity. Both (A2) and (A3) can be weakened. Indeed, there need not exist such a ξ for the method to be used, but, then, a weaker assumption and more detail need to be added. In many problems involving side constraints, $Q(\cdot, \cdot)$ and $A(\cdot)$ are discontinuous and indeed, the function that replaces $A(\cdot)$ may even be multivalued. These matters will be dealt with in a subsequent paper. Also, various abstract valued ξ_n and X_n can be treated. The various possible extensions indicate the power of the technique. Here, we try to keep the structure relatively simple, in order to illustrate the basic ideas.

Both $X^n(\cdot)$ and $I^n(\cdot)$ have paths in $C^r(-\infty, \infty)$ for each n . By (A1) and the boundedness of $Q(\cdot, \cdot)$, both (3.1) and (3.2) hold for $\{X^n(\cdot), I^n(\cdot)\}$. Hence that sequence is tight on $C^{2r}(-\infty, \infty)$.

Henceforth, let N index a weakly convergent subsequence (of the measures induced on $C^{2r}(-\infty, \infty)$ by $\{X^n(\cdot), I^n(\cdot)\}$). The measure on $C^{2r}(-\infty, \infty)$, which is the limit of the weakly convergent subsequence, is induced by a process $X(\cdot), I(\cdot)$ with continuous paths. Using Skorokhod imbedding, $X^N(t) \rightarrow X(t)$, $I^N(t) \rightarrow I(t)$, w.p.1, uniformly on finite t -intervals. The process $I(\cdot)$ is absolutely continuous, and we can suppose that $|\dot{I}(t)| \leq \sup_{x,y} |Q(x,y)|$. Define $\bar{Q}(\cdot)$ by $\dot{I}(t) = \bar{Q}(t)$. Thus, $X(t) = X(0) - \int_0^t \bar{Q}(s) ds$. Of course,

(the measures of) all these limits may conceivably depend on the particular convergent subsequence which is selected. The limit $X(\cdot)$ will now be analyzed, and it will be shown that (the convergence result) (4.5), (4.7), (4.8) hold.

Now, suppose that Skorokhod imbedding is used - so that we can suppose that $\{X^N(\cdot), X(\cdot)\}$ are all defined on the same probability space.

For $t \in (-\infty, \infty)$, define (recall $X^N(s) = X^0(t_N + s)$)

$$E^N(t) = \int_0^t [Q(X^N(s), \bar{\xi}(t_N + s)) - Q(\bar{X}(t_N + s), \bar{\xi}(t_N + s))] ds$$

$$F^N(t) = \int_0^t [Q(X^N(s), \bar{\xi}(t_N + s)) - Q(X(s), \bar{\xi}(t_N + s))] ds.$$

Since $Q(\cdot, \cdot)$ is bounded, $\{E^N(\cdot), F^N(\cdot)\}$ is tight on $C^{2r}(-\infty, \infty)$.

Let N' denote a weakly convergent subsequence of $\{E^N(\cdot), F^N(\cdot), X^N(\cdot), I^N(\cdot), G^N(\cdot)\}$, and suppose (henceforth) that Skorokhod imbedding is used. Note that, by⁺ (A2),

$$|E^{N'}(t)| \leq \max_{|s| \leq T} \theta(|X^{N'}(s) - \bar{X}(t_N + s)|) G^{N'}(t), \quad |t| \leq T.$$

Since $G^{N'}(\cdot)$ converges to some limit $G(\cdot) \in C(-\infty, \infty)$, and the max term goes to zero as $N' \rightarrow \infty$, the limit of $\{E^{N'}(\cdot)\}$ is the zero process. Similarly for $\{F^{N'}(\cdot)\}$. These limits do not depend

⁺The only difference between $X^N(s)$ and $\bar{X}(t_n + s)$ is that the first is a piecewise linear and the second a piecewise constant interpolation. They are equal at the $s = t_{n+i} - t_n$, all i .

on $\{N'\}$, and the actual limit $G(\cdot)$ is not important (only its existence is important), and so we suppose that $N' = N$, and return to the original subsequence. (The above argument will be implicitly used several times in the sequel - usually when (A2) is appealed to.) Thus, the limits of $\{I^N(\cdot)\}$ are the same as those of $\{\bar{I}^N(\cdot)\}$, where we define

$$\bar{I}^N(t) = \int_0^t Q(X(s), \bar{\xi}(t_N + s)) ds, \quad t \geq -t_N,$$

and

$$\bar{I}^N(t) = \bar{I}^N(-t_N), \quad t \leq -t_N.$$

Clearly, $\bar{I}^N(\cdot)$ is easier to work with than is $I^N(\cdot)$. We will now try to simplify $\bar{I}^N(\cdot)$.

Define the function $X_\Delta(\cdot)$ by $X_\Delta(s) = X(i\Delta)$ on $[i\Delta, i\Delta + \Delta)$, $i = 0, \pm 1, \dots$. Define $\bar{I}_\Delta^N(\cdot)$ as $\bar{I}^N(\cdot)$ was defined, but with $X_\Delta(\cdot)$ replacing $X(\cdot)$. Suppose that $\bar{I}_\Delta^N(t) \rightarrow \int_0^t A(X_\Delta(s)) ds$ weakly as $N \rightarrow \infty$. Then by (A2), and the convergence of $X_\Delta(\cdot)$ to $X(\cdot)$ as $\Delta \rightarrow 0$ (uniformly⁺ w.p.1 on finite time intervals), the limit of $\{\bar{I}^N(\cdot)\}$ has values $\int_0^t A(X(s)) ds$. In particular, we can consider one interval at a time: we need only consider the limit of

$$\left\{ \int_{i\Delta}^t Q(x, \bar{\xi}(t_N + s)) ds, \quad t \in [i\Delta, i\Delta + \Delta) \right\} \quad \text{for each } x \in R^r.$$

⁺ Assuming Skorokhod imbedding is used.

Define $\hat{Q}(x, \bar{\xi}(t_N+s)) = Q(x, \bar{\xi}(t_N+s)) - EQ(x, \xi)$. The sequence $\{K^N(\cdot)\}$ with values $(t \geq i\Delta)$

$$K^N(t) = \int_{i\Delta}^t \hat{Q}(x, \bar{\xi}(t_N+s)) ds$$

is tight on $C^r[i\Delta, i\Delta+\Delta]$, since \hat{Q} is bounded. To show $K^N(\cdot) \rightarrow$ zero process weakly, as $N \rightarrow \infty$, it is only necessary to show that $E|K^N(t)|^2 \rightarrow 0$ as $N \rightarrow \infty$. To do this with simple notation, do it component by component. In particular (w.l.o.g.), suppose that $Q(\cdot, \cdot)$ and $\hat{Q}(\cdot, \cdot)$ are scalar valued. Note that $E|K^N(t)|^2$ has the same limit as has

$$(4.4) \quad E \sum_{k, j=m(t_N+i\Delta)}^{m(t_N+t)} a_k a_j \hat{Q}(x, \xi_k) \hat{Q}(x, \xi_j)$$

$$= \sum_{k, j=0}^{m(t_N+t)-m(t_N+i\Delta)} E a_{m(t_N+i\Delta)+k} a_{m(t_N+i\Delta)+j} \hat{Q}(x, \xi_{m(t_N+i\Delta)+k}) \hat{Q}(x, \xi_{m(t_N+i\Delta)+j}).$$

For a $\delta > 0$, $\delta < t - i\Delta$, first consider the sum in (4.4) over those k, j such that $|t_{m(t_N+i\Delta)+k} - t_{m(t_N+i\Delta)+j}| \leq \delta$. This sum is bounded from above by some constant times δ^2 . Thus, to show that (4.4) $\rightarrow 0$ as $N \rightarrow \infty$, we can suppose that $|t_{m(t_N+i\Delta)+k} - t_{m(t_N+i\Delta)+j}| \geq \delta$, for an arbitrary $\delta > 0$. In particular, we can assume that $t_{m(t_N+i\Delta)+k} \geq t_{m(t_N+i\Delta)+j} + \delta$. As $N \rightarrow \infty$, the k, j satisfying this relationship also satisfy $k - j \rightarrow \infty$, since $a_n \rightarrow 0$ as $n \rightarrow \infty$. Using these facts, together with (A3) and the definition of $Q(\cdot, \cdot)$, yields that, for such k, j pairs,

$$E[\hat{Q}(x, \xi_{m(t_N+i\Delta)+k}) | \xi_u, u \leq m(t_N+i\Delta) + j] \rightarrow 0 \text{ w.p.1 as } N \rightarrow \infty.$$

This result, together with

$$\lim_{N \rightarrow \infty} \sum_{j=m(t_N+i\Delta)}^{m(t_N+t)} a_j = (t-i\Delta)$$

and the arbitrariness of δ , implies that (4.4) $\rightarrow 0$ as $N \rightarrow \infty$.

$$\text{Thus, } \bar{I}_\Delta^N(t) \rightarrow \int_0^t A(X_\Delta(s)) ds, \text{ as } N \rightarrow \infty.$$

Finally, combining the above results, we have that

$$\bar{I}^N(t) \rightarrow \int_0^t A(X(s)) ds, \text{ w.p.1, on finite } t\text{-intervals, and, hence,}$$

$$(4.5) \quad \dot{X}(t) = -A(X(t)), \quad t \in (-\infty, \infty),$$

where $X(0)$ may possibly depend on the particular convergent subsequence.

By (A1) and the weak convergence

$$(4.6) \quad \sup_{|t| < \infty} P\{|X(t)| \leq K\} \rightarrow 1 \text{ as } K \rightarrow \infty.$$

By (A4) and (4.6), the paths of $X(\cdot)$ are bounded w.p.1, whatever the convergent subsequence. Under (A4), the bounded trajectories on $(-\infty, \infty)$ must lie in S . Thus, $X(t) \in S, t \in (-\infty, \infty)$. Since S does not depend on the selected subsequence, and since any subsequence of $\{X^N(\cdot), I^N(\cdot)\}$ has a weakly convergent subsequence,

$X_n \rightarrow S$ in probability, as $n \rightarrow \infty$. To strengthen the result, fix T and note that the functions $f(\cdot): C^r(-\infty, \infty) \rightarrow R$ defined by

$$f(x(\cdot)) = \sup_{-T \leq t \leq T} |x(t)|, \text{ or by } \sup_{|t| \leq T} \text{dist.}(x(t), S)$$

are continuous on $C^r(-\infty, \infty)$. Thus, by the weak convergence,

$$(4.7) \quad P\left\{ \sup_{|t| \leq T} \text{dist.}(X^n(t), S) \geq \epsilon \right\} \rightarrow 0$$

as $n \rightarrow \infty$, each $T > 0$, $\epsilon > 0$. It is also true that (for any convergent subsequence N and limit $X(\cdot)$)

$$(4.8) \quad P\left\{ \sup_{|t| \leq T} |X^N(t) - X(t)| \geq \epsilon \right\} \rightarrow 0$$

as $N \rightarrow \infty$, each $\epsilon > 0$.

Unbounded Q. The basic problem concerns tightness of $\{I^N(\cdot)\}$, and some condition which guarantees this need to be added.

A special case is $Q(x, y) = Q_0(x, y) + Q_1(y)$, where Q_0 is bounded, and the processes whose values are the "natural" integrals

$\int_0^t Q_1(\bar{\xi}(t_n + s)) ds$ are tight. Another special case is where

$|Q(x, y)| \leq \text{const.}(1 + |y|)$, and (the constant is independent of n)

$$\int_0^t \int_0^t EQ(\bar{X}(t_n + s), \bar{\xi}(t_n + s)) Q(\bar{X}(t_n + \tau), \bar{\xi}(t_n + \tau)) ds d\tau \leq \text{const.}(t^2)$$

But the matter will not be pursued. The aim here is to illustrate the technique - not to develop the very best conditions.

Remark. The basic convergence result was obtained with relatively little pain, and the proof followed a rather natural set of ideas. The asymptotic problem was reduced to a natural one concerning the properties of an ordinary differential equation (4.5), as in [5], [6], [10]. Many variations are possible. With a suitable alteration in the algorithm, even the case where there is an occasional system disturbance of an impulsive type can be treated, as can various cases where the $\{a_n\}$ are random variables.

5. Inputs $\{\xi_n\}$ Depending on the Iterates $\{X_n\}$

In some applications, it may be desired to let ξ_n depend on X_n , or even on X_n, X_{n-1}, \dots ; for example, in an identification and control problem, where the current input may be allowed to depend on the current parameter estimate. This problem was treated in [5], where a specific application is given, and we re-do the general type of result given there, using the weak convergence technique, and somewhat weaker assumptions on the noise. Let $\{\psi_n\}$ denote a given sequence of random variables, and $h(\cdot, \cdot, \cdot)$ a measurable function such that $\xi_{n+1} = h(\xi_n, \psi_n, X_{n+1})$. We take this form in order to be specific - but more general forms can be used. Additional conditions will be introduced below. Unless otherwise specified, we retain the conditions and terminology of the previous section. Again, our interest is in the techniques and methods of utilizing the weak convergence ideas for the type of general application with which we deal, but not in the more specific applications, of which there are many.

Under (A1) and the bound on $Q(\cdot, \cdot)$, $\{X^n(\cdot), I^n(\cdot)\}$ is still tight on $C^{2r}(-\infty, \infty)$. Fix a weakly convergent subsequence - also to be indexed by N , and with limit $X(\cdot), I(\cdot)$. Note that (4.8) still holds. To get a convergence result, the net or average effect of X_n on ξ_m must vanish as $m - n \rightarrow \infty$ and some condition guaranteeing this must be introduced. To this end and for reference below let $q_n \rightarrow \infty$ as $n \rightarrow \infty$ and define, for each x and $n = 0, \dots$, the sequences

$$(5.1) \quad \begin{aligned} \bar{\xi}_{k+1}^n(x) &= h(\bar{\xi}_k^n(x), \psi_{Q_{n+k}, x}), \\ \hat{\xi}_{k+1}^n(x) &= h(\hat{\xi}_k^n(x), \psi_{Q_{n+k}, x}) + \varepsilon_k^n, \quad k = 0, 1, \dots, \end{aligned}$$

where, for each x and n , $\hat{\xi}_0^n(x) = \bar{\xi}_0^n(x)$, and $\{\varepsilon_k^n, k = 0, 1, \dots\}$ is a sequence of "errors".

Assume

(A5) $\{\xi_n\}$ is tight, or, equivalently, bounded in probability, uniformly in, (on the Euclidean range space); i.e.,

$$\sup_n P\{|\xi_n| \geq K\} \rightarrow 0 \quad \text{as} \quad K \rightarrow \infty.$$

(A6) (replaces (A3)) If, for some fixed x , $\{\bar{\xi}_0^n(x)\}$ is tight⁺, then there is a random variable ξ^x such that, for each bounded and continuous function $f(\cdot)$,

$$E[f(\bar{\xi}_{i+k}^n(x)) | \bar{\xi}_j^n(x), j \leq i] \rightarrow Ef(\xi^x),$$

uniformly in n , as $i, k \rightarrow \infty$. The function of (x, y) with values $EQ(y, \xi^x)$ is continuous. Define $A(x) = EQ(x, \xi^x)$, and let $A(\cdot)$ have the properties of the $A(\cdot)$ in (A4).

(A7) For each $\delta > 0$, $t > 0$, suppose that there is an $\varepsilon > 0$ such that $\{|\varepsilon_k^n| \leq \varepsilon \text{ for } 0 \leq k \leq m(t_n + t) - n\}$ and tightness⁺ of the initial sequence $\{\bar{\xi}_0^n(x)\} = \{\hat{\xi}_0^n(x)\}$ imply that

$$\lim_{n \rightarrow \infty} \sup_{k: k \leq m(t_n + t) - n} P\{|\bar{\xi}_k^n(x) - \hat{\xi}_k^n(x)| \geq \delta\} \leq \delta.$$

⁺or, equivalently, boundedness in probability, uniformly in n .

Assumptions (A5), and (A6) do not seem to be restrictive, and the condition requiring existence of ξ^x can also be weakened. Condition (A7) says, basically, that the $\{\xi_n\}$ process has an inherent stability for each fixed $X_n \equiv x$, in the sense that small perturbations to $h(\cdot, \cdot, \cdot)$ do not seriously affect the value of $\{\xi_n\}$.

(A8) $h(\cdot, \cdot, \cdot)$ is continuous in its third argument, uniformly in the first two: I.e., for some $\theta(\cdot)$ with $\theta(u) \rightarrow 0$ as $u \rightarrow 0$,

$$|h(\xi, \psi, x') - h(\xi, \psi, x)| \leq \theta(|x-x'|), \text{ uniformly in } \xi, \psi.$$

Condition (A8) can be replaced by more general alternatives. For example, if (A7) is altered such that $\{|\epsilon_k^n| \leq \epsilon, \dots\}$ is replaced by $\{\sup_{i \leq m(t_n+t)-n} \sum_{k=0}^i |\epsilon_k^n| \leq \epsilon\}$, then (A8) can be replaced by

(A8') Let there exist $\theta(\cdot), g_1(\cdot)$ such that $\theta(u) \rightarrow 0$ as $u \rightarrow 0$ and

$$|h(\xi, \psi, x') - h(\xi, \psi, x)| \leq \theta(|x-x'|)g_1(\xi, \psi),$$

and let $\{G_1^n(t)\}$ be tight $(G_1^n(\cdot))$ is defined analogously to $G^n(\cdot)$ in (A2).

The aim of this Section is to show the same end result as was shown in Section 4, namely that $X(\cdot)$ satisfies $\dot{X}(t) = -A(X(t))$ ($A(\cdot)$ defined in (A6)), from which we get (as in Section 4) that $X(t) \xrightarrow{P} S$, and (4.7). As in Section 4, it suffices to show that

$$(5.2) \quad \int_{i\Delta}^t Q(X(i\Delta), \bar{\xi}(t_N+s)) ds \rightarrow \int_{i\Delta}^t EQ(x, \xi^x) ds \Big|_{x=X(i\Delta)}$$

weakly on $C^r[i\Delta, i\Delta+\Delta]$ as $N \rightarrow \infty$, for each $\Delta > 0$ and i . We always suppose that Skorokhod imbedding is used. Now, define the initial condition and error terms in (5.1) by (we select them to enable us to deal with the processes on the interval $[i\Delta, i\Delta+\Delta]$)

$$\begin{aligned} \hat{\xi}_0^N(x) &= \xi_m(t_N+i\Delta) = \bar{\xi}_0^N(x), \text{ all } x, \\ q_N &= m(t_N+i\Delta), \end{aligned}$$

$$\begin{aligned} \varepsilon_k^N &= h(\xi_m(t_N+i\Delta)+k, \psi_m(t_N+i\Delta)+k, X_m(t_N+i\Delta)+k+1) \\ &\quad - h(\xi_m(t_N+i\Delta)+k, \psi_m(t_N+i\Delta)+k, X(i\Delta)). \end{aligned}$$

Thus

$$(5.3a) \quad \xi_m(t_N+i\Delta)+k = \hat{\xi}_k^N(X(i\Delta))$$

and

$$(5.3b) \quad \bar{\xi}(t_N+i\Delta+u) = \hat{\xi}_{m(t_N+i\Delta)+m(t_N+i\Delta)+u}^N(X(i\Delta)), \quad \text{for } u \geq 0.$$

By (A5), the tightness requirement on $\{\hat{\xi}_0^N(X(i\Delta))\} = \{\xi_m(t_N+i\Delta)\}$ in (A7) (for $x = X(i\Delta)$) is satisfied. Next, note that (5.4) follows from (A8) and (4.8).

$$(5.4) \quad \lim_{\Delta \rightarrow 0} \lim_{N \rightarrow \infty} P\left\{ \sup_{0 \leq k \leq m(t_N + i\Delta + \Delta) - m(t_N + i\Delta)} |\xi_k^N| \geq \varepsilon \right\} = 0, \text{ each } \varepsilon > 0.$$

By (5.3), (5.4) and (A7), there is a δ_1 which goes to 0 as $\Delta \rightarrow 0$ and such that

$$(5.5) \quad \lim_{N \rightarrow \infty} E \int_{i\Delta}^{i\Delta + \Delta} |Q(X(i\Delta), \bar{\xi}(t_N + s)) - Q(X(i\Delta), \bar{\xi}^N(X(i\Delta), s))| ds \leq \delta_1 \Delta,$$

where we define $\bar{\xi}^N(x, s)$ to be the piecewise constant right continuous interpolation of $\{\xi_k^N(x)\}$, with interpolation intervals

$$\{a_{m(t_N + i\Delta)}, a_{m(t_N + i\Delta) + 1}, \dots\}$$

By (5.5), in order to prove (5.2), we need only show that the sequence of functions $\hat{Q}^N(\cdot)$ on the interval $[i\Delta, i\Delta + \Delta]$ and with values

$$(5.6) \quad \hat{Q}^N(t) = \int_{i\Delta}^t \hat{Q}(x, \bar{\xi}^N(x, s)) ds,$$

where

$$\hat{Q}(x, \bar{\xi}^N(x, s)) = Q(x, \bar{\xi}^N(x, s)) - EQ(x, \xi^x),$$

tends weakly to the zero process as $N \rightarrow \infty$, for each x . As done in Section 4, we can suppose that $Q(\cdot, \cdot)$ is scalar valued. Then,

(5.6) has the value

$$(5.7) \quad \sum_{j,k=0}^{m(t_N+i\Delta+t) - m(t_N+i\Delta)} [a_{m(t_N+i\Delta)+k} a_{m(t_N+i\Delta)+j}] \hat{Q}(x, \bar{\xi}_k^N(x)) \hat{Q}(x, \bar{\xi}_j^N(x)).$$

Now, by (A6), and an argument like that associated with the convergence of (4.4), and with $f(\cdot) = \hat{Q}(x, \cdot)$, we have $E|\hat{Q}^N(t)|^2 \rightarrow 0$, which implies that $\hat{Q}^N(\cdot) \rightarrow$ zero process weakly. The demonstration of convergence is now complete, and (4.5) and (4.7) continue to hold.

The proof was relatively straightforward, and the assumptions not unreasonable. The tightness assumptions and weak convergence techniques allow a relatively simple treatment - one which focuses on the basic structures and does not get overinvolved in detail. Generalizations to abstract valued problems are also possible.

6. An Identification Problem

Again, following an application in [5], we discuss an algorithm for the identification of the coefficients $\theta = (A_1, \dots, A_\ell, \dots, B_1, \dots, B_\ell)$ in the system

$$(6.1) \quad Y_n + A_1 Y_{n-1} + \dots + A_\ell Y_{n-\ell} = B_1 u_{n-1} + \dots + B_\ell u_{n-\ell} + \rho_n,$$

where $\{\rho_n\}$ is some sequence of random variables. For notational simplicity, let the A_i, B_j be scalars; the general case is treated in exactly the same way. Define $\psi_n = (-Y_{n-1}, \dots, -Y_{n-\ell}, u_{n-1}, \dots, u_{n-\ell})$. Then $Y_n = \theta' \psi_n + \rho_n$. Suppose that the l.h.s. of (6.1) is asymptotically stable, and $\{u_n, \rho_n\}$ satisfies the condition in (A3) on $\{\xi_n\}$. Let ψ_n be known at time n .

Let Y_n denote the n^{th} estimate of θ . A common recursive estimation algorithm is given by (6.1) - for suitable functions $Q_1(\cdot), Q_2(\cdot)$.

$$(6.1) \quad \begin{aligned} R_n &= R_{n-1} + a_n Q_1(\psi_n \psi_n' - R_{n-1}) \\ K_n &= R_n^{-1} \psi_n / [1 + a_n (\psi_n' R_n^{-1} \psi_n - 1)] \\ Y_n &= Y_{n-1} + a_n Q_2(K_n (Y_n - Y_{n-1}' \psi_n)) \end{aligned}$$

Usually the $Q_i(\cdot)$ are the identity functions. Let $Q_1(\cdot) =$ identity, and suppose that

$$(6.2) \quad R_n, Y_n \quad \underline{\text{is bounded in probability, uniformly in } n},$$

and that there is a positive definite matrix H such that

$$(6.3) \quad E[\psi_{n+k} \psi_{n+k}' | \psi_i, i \leq n] \rightarrow H, \text{ as } n, k \rightarrow \infty.$$

Define $B^0(\cdot)$: $B^0(t) = 0, t \leq 0$, and $B^0(t) = B^0(t_n) + (t-t_n)\psi_n\psi_n'$ on $[t_{n+1} - t_n]$, and define $B^n(\cdot)$ by $B^n(t) = B^0(t+t_n)$. Also, define $R^n(t)$, and $Y^n(t)$ as $X^n(\cdot)$ was defined in Section 4, but with the r.h.s. of the first and third lines of (6.1) replacing the r.h.s. of (4.2). Only a rough outline will be given. Under reasonable conditions on $\{\psi_n\}$, the sequence $\{B^n(\cdot), R^n(\cdot)\}$ is tight on $^{\dagger} C^{2s}(-\infty, \infty)$. Then, by the result of Section 4, $\{R_n\}$ tends in probability to the constant limit solution \bar{R} of

$$\dot{R} = H - R,$$

and (4.7) holds with $X^N(t)$, S replaced by $R^n(t), \bar{R}$.

The convergence of $\{Y_n\}$ can also be treated with $Q_2 = \text{identity}$. But in order to be able to appeal directly to the result of Section 4, without further work, let $Q_2(\cdot)$ be bounded and continuous: for example, each component of $Q_2(\cdot)$ can be a saturation function. (Indeed, it is possible to study the limit as a function of the saturation level.) Some additional conditions need to be introduced to assure that (A2) and (A3) hold (here ξ_n is replaced by ψ_n, ρ_n). These conditions are not unreasonable, but to save some space and discussion, simply assume (A2), (A3). Define $A_2(\cdot, \cdot)$ (a function of a matrix M and vector Y) by

$$A_2(M, Y) = \lim_{n, k \rightarrow \infty} E Q_2(M \psi_{n+k} (Y_{n+k} - Y' \psi_{n+k}) | Y_i, \psi_i, i \leq n).$$

[†] s = number of elements in R_n .

Let $A_2(\cdot, \cdot)$ and $\dot{Y} = A_2(\bar{R}^{-1}, Y)$ have the properties of the $A(\cdot)$ and $\dot{x} = -A(x)$ of (A4). Then Y_n converges in probability to the limit set of $\dot{Y} = A_2(\bar{R}^{-1}, Y)$ and (4.7) holds also.

7. Conclusions

Some of the concepts of weak convergence theory have been introduced and applied to convergence problems for a family of recursive adaptive procedures. The conditions and ideas are rather natural for that type of problem, and the proofs are relatively simple. There are possible extensions in many directions. It is expected that the techniques will play an important role in control theory.

References

- [1]. D.E. Iglehart, "Diffusion approximations in applied probability", Mathematics of the Decision Sciences, Part II, Lectures in Applied Math., 12, 1968, American Math. Soc., Providence, R.I.
- [2] D.E. Iglehart, "Weak convergence in queueing theory", J. Appl. Prob., 5, pp. 570-594, 1973.
- [3] H.J. Kushner, A survey of some applications of probability and stochastic control theory to finite difference methods for degenerate elliptic and parabolic equations", SIAM Rev., 18, no. 4, pp. 545-577, 1976.
- [4] H.J. Kushner, Probability Methods for Approximations in Stochastic Control and for Elliptic Equations, Academic Press, to appear February, 1977.
- [5] L. Ljung, "Analysis of recursive stochastic algorithms", Report 7616(C) March 1976, Dept. of Automatic Control, Lund Inst. of Technology. Submitted to IEEE Trans. on Automatic Control.
- [6] L. Ljung, T. Söderström, I. Gustavsson, "Counterexamples to general convergence of a commonly used recursive identification method", IEEE Trans. on Automatic Control, AC-20, no. 5, October 1975, pp. 643-652.
- [7] P. Billingsley, Convergence of Probability Measures. John Wiley, New York, 1968.
- [8] I.I. Gikhman, A.V. Skorokhod, Introduction to the Theory of Random Processes, Saunders, Phila., 1969.
- [9] A.V. Skorokhod, "Limit theorems for stochastic processes", Theory of Probability and its Applications, 1, 1956, pp. 262-290 (English translation).
- [10] H.J. Kushner, "General convergence results for constrained and unconstrained stochastic approximations", 1976 Decision and Control meeting, St. Petersburg, Florida.

<p>18 19 REPORT DOCUMENTATION PAGE</p>		<p>READ INSTRUCTIONS BEFORE COMPLETING FORM</p>	
<p>1. REPORT NUMBER AFOSR - TR - 76 - 1410</p>		<p>2. GOVT ACCESSION NO. 3. RECIPIENT'S CATALOG NUMBER</p>	
<p>4. TITLE (and Subtitle) CONVERGENCE OF RECURSIVE ADAPTIVE AND IDENTIFICATION PROCEDURES VIA WEAK CONVERGENCE THEORY,</p>		<p>5. TYPE OF REPORT & PERIOD COVERED 9 Interim rept.</p>	
<p>7. AUTHOR(s) 10 Harold J. Kushner</p>		<p>15 8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0279 AF-AFOSR 76-3063-76</p>	
<p>9. PERFORMING ORGANIZATION NAME AND ADDRESS Brown University, Lefschetz Ctr for Dynamical Systems, Div of Applied Mathematics Providence, RI 02912</p>		<p>10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 12304/A1</p>	
<p>11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332</p>		<p>12. REPORT DATE 11 1976 17 A1</p>	
<p>14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 1234p.</p>		<p>13. NUMBER OF PAGES 31</p>	
		<p>15. SECURITY CLASS. (of this report) UNCLASSIFIED</p>	
		<p>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE</p>	
<p>16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.</p>			
<p>17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)</p>			
<p>18. SUPPLEMENTARY NOTES</p>			
<p>19. KEY WORDS (Continue on reverse side if necessary and identify by block number)</p>			
<p>20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Results and concepts in the theory of weak convergence of a sequence of probability measures are applied to convergence problems for a variety of recursive adaptive (stochastic approximation like) methods. Similar techniques have had wide applicability in areas of operations research and in some other areas in stochastic control. It is quite likely that they will play a much more important role in control theory than they do at present, since →</p>			

20 Abstract

cont

→ they allow relatively simple and natural proofs for many types of convergence and approximation problems. Part of the aim of the paper is tutorial: to introduce the ideas, and to show how they might be applied. Also, many of the results are new, and they can all be generalized in many directions.



