City University of New York (CUNY)

CUNY Academic Works

Publications and Research

Hunter College

1996

Stochastic Approximation Methods for Systems Over an Infinite Horizon

Harold J. Kushner
Brown University

Felisa Vázquez-Abad
CUNY Hunter College

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/hc_pubs/62

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

STOCHASTIC APPROXIMATION METHODS FOR SYSTEMS OVER AN INFINITE HORIZON*

HAROLD J. KUSHNER[†] AND FELISA J. VÁZQUEZ-ABAD[‡]

Abstract. The paper develops efficient and general stochastic approximation (SA) methods for improving the operation of parametrized systems of either the continuous- or discrete-event dynamical systems types and which are of interest over a long time period. For example, one might wish to optimize or improve the stationary (or average cost per unit time) performance by adjusting the systems parameters. The number of applications and the associated literature are increasing at a rapid rate. This is partly due to the increasing activity in computing pathwise derivatives and adapting them to the average-cost problem. Although the original motivation and the examples come from an interest in the infinite-horizon problem, the techniques and results are of general applicability in SA. We present an updating and review of powerful ordinary differential equation-type methods, in a fairly general context, and based on weak convergence ideas. The results and proof techniques are applicable to a wide variety of applications. Exploiting the full potential of these ideas can greatly simplify and extend much current work. Their breadth as well as the relative ease of using the basic ideas are illustrated in detail via typical examples drawn from discrete-event dynamical systems, piecewise deterministic dynamical systems, and a stochastic differential equations model. In these particular illustrations, we use either infinitesimal perturbation analysis-type estimators, mean square derivative-type estimators, or finite-difference type estimators. Markov and non-Markov models are discussed. The algorithms for distributed/asynchronous updating as well as the fully synchronous schemes are developed.

Key words. stochastic approximation, ordinary differential equation method, weak convergence, recursive optimization, Monte Carlo optimization, discrete-event dynamical systems, piecewise deterministic dynamical systems, stationary cost problems

AMS subject classifications. 62L20, 93C40, 93E25, 90B25

1. Introduction. The paper is concerned with efficient and general stochastic approximation (SA) methods for parametrized systems of either continuous or discrete event dynamical systems that are of interest over a long time period. For example, one might wish to optimize or improve the stationary (or average cost per unit time) performance by adjusting the systems parameters. The number of applications and the associated literature are increasing at a rapid rate. Although the motivation and examples come from an interest in this infinite-horizon problem, the techniques and results are of general applicability in SA. Basic techniques for such problems have appeared in [2, 22, 27]. These techniques are still fundamental for applications to the general problems of current interest. Exploiting their full potential can greatly simplify and extend much current work. We present a full development of the basic ideas in [22, 27] and related works in a more general context, with the particular goal of illustrating their breadth as well as the relative ease of using them in particular applications.

To fix ideas, let θ denote an adjustable parameter of a dynamical system and $x(\cdot, \theta)$ the associated system state process. For a cost rate $c(\theta, x)$, define $C_T(\theta, x(0)) = E \int_0^T c(\theta, x(t, \theta)) dt / T$ and $C(\theta, x(0)) = \lim C_T(\theta, x(0))$. We wish to minimize $C(\theta, x(0))$

* Received by the editors February 23, 1994; accepted for publication (in revised form) May 12, 1995.

[†] Applied Mathematics Department, Brown University, Providence, RI 02912. The research of this author was supported by NSF grant ECS-9302137, AFOSR contract F49620-92-J-0081, and ARO contract DAAL03-92-G-01157.

[‡] Département d'Informatique et Recherche Opérationnelle, Université de Montréal, Montréal, PQ, H3C 3J7, Canada. The research of this author was supported by NSERC grant WFA0139015.

by the dynamic adjustment of the parameter θ , using estimates of the derivatives made from measurements of the sample path. Indeed, much of the recent interest in SA methods has been motivated by the increasing availability of good estimators of the derivatives of objects such as $C_T(\theta, x(0))$, say, of the infinitesimal perturbation analysis (IPA) or related types [13, 14, 18, 34, 42, 45] or of the mean square derivative type [5]. With ϵ_n a step-size parameter and θ_n the n th estimate of the parameter, the basic SA algorithm is $\theta_{n+1} = \theta_n + \epsilon_n Y_n$, where Y_n is the measurement used for the current update. One is concerned with the asymptotic properties of the sequence θ_n . The ordinary differential equations (ODE) method shows that the asymptotic properties can be characterized in terms of the limit properties of the solution to an ODE $\dot{\theta} = g(\theta)$, where, loosely speaking, $g(\theta)$ is the stationary mean value of Y_n given that the parameter value is always fixed at θ . Thus the individual Y_n themselves need not be (asymptotically) unbiased estimators of the gradient at the current parameter values. The fact that the estimators are taken over a finite time interval but one actually wishes to use them effectively for the infinite-time problem has led to various ad hoc approaches, often driven by the proof technique. One technique was to let the successive estimation intervals go to infinity. It will appear from the results in §§3–5 (a direct consequence of the results in [22, 27]) that to get the desired limit result one generally need not reinitialize the estimator periodically nor let the intervals go to infinity. One basically does what is more natural: keep the successive updating time intervals bounded and appropriately update the estimator without “reinitializing” it. The proofs of such results are the essence of the “local averaging” intuition in the ODE method, initiated by Ljung [33], although the techniques used here are quite different.

The paper is not concerned with optimization per se but rather with getting the appropriate ODE for the SA algorithm of interest and in showing the great flexibility in the algorithms that one can use and analyze. For the optimization problem, one generally needs to show that the solution of the ODE converges to the desired point, and this requires a closer look at the right-hand side of the ODE. In some cases, this involves showing that the right side of the ODE is the negative of the gradient of a desired cost function with a particular structure. Indeed, in §§7 and 8, we show that the right side is indeed the negative of the gradient of the desired ergodic cost. But in any application, one needs first to characterize the correct ODE and then to analyze the limits of its solutions. The latter job is highly problem dependent.

One can try to prove that the convergence either is with probability one (w.p.1) or is in a weak (or generalized distributional) sense. Our framework for getting the asymptotic properties is that of weak convergence. This allows the use of what might be the simplest mathematical techniques and conditions. For example, for the SA with decreasing step sizes $0 < \epsilon_n \rightarrow 0$ satisfying $\sum \epsilon_n = \infty$, no additional conditions need be imposed on the ϵ_n . Conditions of the often used type [2] $\sum \epsilon_n^{1+\alpha} < \infty$ for some $\alpha > 0$ are not needed. The sequence of estimators need only be uniformly integrable, and no additional moment conditions are needed. The weak convergence technique correctly identifies the places where the process spends either almost all or all of its (asymptotic) time, and gives us a fairly complete stability structure of the algorithm.

For the decreasing step-size algorithms, the difference between the probability-one and the weak convergence results is not as great as what one might at first suppose. Indeed, known results show that under quite weak additional conditions, probability one convergence follows directly from the weak convergence results, and we

now comment loosely on this. Suppose that the ODE is locally asymptotically stable about a point θ^* with open domain of attraction Γ . The ODE method associated with the weak convergence approach quite generally allows us to show that some such set Γ is entered infinitely often. Then, under very weak conditions, one can appeal to existing applications of large deviations methods to SA's to get probability one convergence. This idea is fully developed in [9]. Among other things, it is shown in this reference that one gets probability one convergence with the only additional requirement on the step-size sequence ϵ_n is that it satisfy

$$\sum_n e^{-\delta/\epsilon_n} < \infty$$

for each $\delta > 0$. That is, we need only that $\epsilon_n < c_n/\log n$, where $c_n \rightarrow 0$. The conditions on the noise process in [9] are satisfied by the usual processes that are not too "heavily tailed," and for such processes these probability one convergence results might be about the best now available. The main point is that once the weak convergence results are available, probability one results follow directly from existing works under broad conditions, and the basic weak convergence techniques are very much simpler than those required for probability-one convergence.

It is worth noting that in applications, probability-one results might be of illusory advantage over weak convergence results. The algorithms generally have stopping rules and, when these are applied, one generally has only probabilistic or distributional information about the last iterate.

We are also concerned with constant step size cases where one can only use weak (and not w.p.1) convergence ideas. Indeed, in problems of tracking time-varying systems one must use constant step sizes. In adaptive problems in communication theory and signal processing, constant step sizes are the common practice. Even if the problem is such that decreasing step sizes can be used, one often lets them be constant due to robustness considerations. Indeed, in practice one often prefers algorithms which get to a neighborhood of the desired point quickly, and this argues for a constant step size.

The development in the paper requires only some of the elementary concepts from the theory of weak convergence. These are reviewed in §2. Perhaps the only required nonelementary fact concerns the use in Theorem 3.1 of random variables which are measure valued. Our application of this concept is straightforward, since for our purposes the important facts concerning such random variables are determined by their mean values and will be implied by the conditions imposed on the "noise" terms. The concept of measure-valued random variables allows us to deal more easily than in the past with unbounded noise.

The basic result of the paper is Theorem 3.1. It is basic in that it lays out the fundamental ideas of the averaging method, and most subsequent results can be derived by mild modifications of the technique of that theorem. The theorem is for the constant-step-size case. But, as seen in §4, the case where $\epsilon_n \rightarrow 0$ differs only in the way certain terms are grouped in the proof. In Theorem 3.1, we have tried to use conditions that are fairly general. Since one's imagination in constructing algorithms is endless, no set of conditions is "completely general." But it will be seen that the conditions used are quite minimal, and allow the few basic ideas to be exposed. The first basic idea is to repose the problem as a "martingale problem," which allows us to replace the noise terms by appropriate conditional expectations given the past, and greatly facilitates the averaging. Then we are confronted by the fact that the

noise at step n can depend on the values of the state at that time as well as at previous times. In Theorem 3.1, this is handled in a convenient way (coming originally from [27]) by the use of a Markov model for the joint (noise, state) process, and imposing appropriate weak continuity conditions on the transition function. (Non-Markov models are treated in Appendix 1, but the Markov assumption in Theorem 3.1 is quite powerful, since the state space can be a complete separable metric space, thus allowing convenient “Markovianizations.”) In doing the local averaging to get the appropriate ODE, these weak continuity assumptions allow us to average as though the state did not change. They facilitate the use of the appropriate (mean) ergodic theorems for the noise processes which, for the purposes of averaging, can be assumed to evolve as though the state did not change. These few basic and powerful ideas underlie all the results and are widely adaptable. The averaging idea of Theorem 3.1 is like that in [27] but is somewhat more general, particularly in the treatment of unbounded noise.

Section 3.2 concerns the asymptotic points of the algorithm, and in Theorem 3.2 they are identified with the two-sided invariant set (in the sense of dynamical systems theory) of the ODE. The other parts of §3 concern the simplifications when the basic observation has an “additive” character or the problem has a regenerative structure and one wishes to update at the regeneration times. This “additivity” property is common to numerous applications, as seen in §§6–9. In general, updating at regenerative intervals, even if the process has a regenerative structure, is not needed and might not even be a good idea. More will be said about this later. It is certainly inadvisable when the regenerative periods are very long. In §4, we make the few necessary changes when the step sizes ϵ_n go to zero.

Section 5 gives the simple alterations when the iterate is to be confined to some constraint set. It was noted in [25] and elsewhere subsequently that the ODE for the constrained problem follows directly from that for the unconstrained problem by use of a simple decomposition of the iterate into the sum of the unprojected value plus an “error.” The “error” is easy to treat since it is what brings an infeasible point back to the constraint set. The unprojected values are treated as for the unconstrained algorithm. So, under appropriate conditions on the constraint set, the constrained problems are easy extensions of the unconstrained problem.

Section 6 formally introduces the application of Theorem 3.1 and its extensions for use on systems whose performance function involves a stationary average. The basic heuristic illustration is for a system where an IPA- or mean square derivative-type estimator might be used and we wish to minimize a stationary cost. The right side of the limit ODE is the negative of the derivative of the stationary cost with respect to the adjustable parameter. All of this is a consequence of the basic theorem. Many authors [6, 7, 16, 32, 34, 45] consider finite-horizon gradient estimators. They reset the estimation (reset the accumulator, to use current jargon) at the start of each observation interval, whose length becomes large as $n \rightarrow \infty$. It will be seen that quite often one does not need to let the observation intervals become large nor to reset the estimator. Indeed, these latter techniques are frequently adopted just because it is under those conditions that the authors have proved their convergence results.

To illustrate the basic simplicity and power of the approach, in §§7 and 8 we have chosen examples of current importance and on which much work has been done. Each example is typical of a large class of great current interest and illustrates the application of the methods to that class. The problem of §7 concerns the optimization of a single queue with respect to a service time parameter. This problem has been well

studied and is typical of the use of IPA in many discrete event dynamical systems. The problem in §8 concerns the optimization of an “unreliable” manufacturing system via the choice of suitable production rates and thresholds and is typical of many applications to piecewise deterministic systems. In both cases, the general techniques discussed here are relatively quick to apply and yield good results for many forms of the algorithms and under conditions which are weaker than those generally used. The power of the approach allows much flexibility in the SA algorithm.

In §9 we apply the ideas of §3 to a stochastic differential equation (SDE) model, where the sample derivative is obtained from the equation for the adjoint or mean square derivative. This is just the SDE analogue of the IPA-type estimator and has been in use for a long time. In such examples one often has the problem of proving stability of the derivative estimators, and there is no regenerative structure to help. When stability can be proved, the results are exactly as for the discrete-event and piecewise deterministic dynamical systems cases; one need not restart the estimator nor let the estimation periods increase with time, each of which might not be good practice. In the limit one gets the basic ODE, whose right side is the negative of the gradient of the stationary cost with respect to the parameter. When stability of the mean square derivative process cannot be proved, one can use various forms of finite differences. For example, one might use one continuous run either with the parameter being perturbed over successive intervals of, say, fixed length, or with the use of independent samples for the positive and negative perturbations. In the former (one sample) case, it is noteworthy that we can often get something close to the desired limit ODE. Either finite difference method can be employed when the functions in the cost or dynamical equation are not smooth (say, the cost involves the indicator function of some event) or when we do not know the model well enough to even try to compute a pathwise derivative.

The appendices contain various extensions. Appendix 1 uses a perturbed test function type method (of the type used in [23]) to avoid the Markov assumption. In Appendix 2, we illustrate the use of a method with which one can sometimes avoid the use of occupation measures in the argument of Theorem 3.1 and which is adaptable to many uses. Appendix 3 contains the few additional details when one wishes to work within a regenerative context but possibly update at rather arbitrary random times during the interval as well as at its end. Appendix 4 contains the essential ideas for dealing with a decentralized algorithm, where the different processors update on their own (asynchronous) schedule, with possible delays in communication. Using simple time-change arguments (extensions of the type first used in [30]), we show that the proof and the end results are essentially as for the basic synchronized case, except for some notational changes. This approach generalizes the results in [40]. Thus, the described approach efficiently encompasses a very diverse group of algorithms and applications.

Although the essential ideas are all in Theorem 3.1, the paper is long because we wish to show the great flexibility of the ideas and how to extend them effectively in the many possible (not entirely obvious) directions which are of increasing current interest and to properly illustrate their practical use via concrete applications to important problems.

We note that the convergence can generally be accelerated using the iterate averaging methods initiated by Polyak and discussed in [29, 36, 46, 47].

2. Some background on weak convergence. The methods of the theory of weak convergence are powerful and widely used tools for problems concerning approx-

imations and limit theorems for random processes [3, 10, 22]. They do for random processes what the central limit theorem and the law of large numbers do for sequences of vector-valued random variables. Because they are averaging methods for random processes evolving on different time scales, they are natural methods for SA and have been widely used. Only the basic definitions will be given, since the ideas will be used in a simple way. Further information for those interested can be found in the references.

Let $\{X_n, n < \infty\}$ be a sequence of random variables with values in a complete and separable metric space (CSMS) S . In this paper, S will generally be either some Euclidean space R^k , a space of functions representing the paths of the SA process, or a set of probability measures, as specified below. We say that $\{X_n, n < \infty\}$ *converges weakly* to a random variable X and write $X_n \Rightarrow X$, if for each continuous and bounded real-valued function $f(\cdot)$ on S we have $Ef(X_n) \rightarrow Ef(X)$. Thus, weak convergence is an extension of the concept of convergence in distribution of a sequence of real valued random variables to more general spaces. If P_n and P , resp., are the measures of X_n and X , resp., we also say that $P_n \Rightarrow P$. The sequence $\{X_n, n < \infty\}$ is said to be *tight* if for each $\delta > 0$ there is a compact set $K_\delta \subset S$ such that $P\{X_n \notin K_\delta\} \leq \delta$ for all n . Equivalently, a set of measures $\{P_n, n < \infty\}$ on the Borel sets of S is said to be tight if $P_n\{S - K_\delta\} \leq \delta$ for all n . Tightness implies the existence of a weakly convergent subsequence [10, p. 104].

The Skorohod representation. Since weak convergence is a generalized distributional convergence, it does not depend on the actual probability space that is used. It is often more convenient in the analysis to work with w.p.1 convergence rather than with weak convergence directly. The *Skorohod representation* [10] guarantees that we can choose the probability space so that w.p.1 convergence holds if weak convergence does, as follows. Suppose that $X_n \Rightarrow X$ weakly. Then we can find a probability space with random variables $\{\tilde{X}_n, n < \infty\}$, $\tilde{X}$ defined on it, where $\tilde{X}_n$ (resp., $\tilde{X}$) has the same measure as X_n (resp., X) and on which $\tilde{X}_n \rightarrow \tilde{X}$ w.p.1 in the topology of S [10, p. 102]. We will use the Skorohod representation where convenient.

The path spaces. Define $D^r[0, \infty)$ or $D^r(-\infty, \infty)$, where $D^r(I)$ is the space of R^r -valued functions on the interval I which are right continuous and have left-hand limits (and are continuous at $t = 0$ in the case of $D^r[0, \infty)$). The topology will be that of uniform convergence on finite intervals, making both spaces into CSMSs.

Notation on interpolated processes. In the SA algorithms that we study, we will have recursions of the form $X_{n+1}^\epsilon = X_n^\epsilon + \epsilon \chi_n^\epsilon$, where χ_n^ϵ is a sequence of R^r -valued random variables. We are interested in studying the limit behavior as $\epsilon \rightarrow 0$ and $n \rightarrow \infty$. The piecewise constant interpolation of X_n^ϵ on $(-\infty, \infty)$ is defined as $X^\epsilon(t) = X_n^\epsilon$ for $t \in [\epsilon n, \epsilon(n+1))$ and $X^\epsilon(t) = X_0^\epsilon$ for $t < 0$. For $t \geq 0$, let $[t]$ denote the integer part of t . Then,

$$(2.1) \quad X^\epsilon(t) = X(0) + \epsilon \sum_{i=0}^{[t/\epsilon]-1} \chi_i^\epsilon, \quad t \geq 0.$$

$X^\epsilon(\cdot)$ is a random process with paths in $D^r(-\infty, \infty)$. We also view it as a random variable with values in $D^r(-\infty, \infty)$. We will also use shifted processes, as follows. Let q_ϵ be a sequence of integers such that $\epsilon q_\epsilon \rightarrow \infty$ as $\epsilon \rightarrow 0$. Then $X^\epsilon(\epsilon q_\epsilon + \cdot)$ will also be of interest, since the “tail” of the original process is now in the “vicinity of the origin” for large ϵq_ϵ . We will be interested in the limits of the $X^\epsilon(\cdot)$ as $\epsilon \rightarrow 0$. For notational simplicity, we henceforth write $[t/\epsilon]$ simply as t/ϵ in the limits of the sums.

Next, let ϵ_j be a sequence of positive numbers which goes to zero and such that $\sum_j \epsilon_j = \infty$. Let χ_i be a sequence of R^r -valued random variables. Define $t_n = \sum_{i=0}^{n-1} \epsilon_i$, $m(t) = \max\{i : t_i \leq t\}$. With $X(0)$ given, define the interpolation $X^0(\cdot)$ by $X^0(t) = X(0)$ for $t \leq 0$, and for $t \geq 0$,

$$(2.2) \quad X^0(t) = X(0) + \sum_{i=0}^{m(t)-1} \epsilon_i \chi_i.$$

The shifted processes $X^n(t) = X^0(t + t_n)$ will play an important role since they bring the tail of $X^0(\cdot)$ to the forefront. The following result will be used. It is not hard to prove directly and follows from [3, Thm. 15.2].

LEMMA 2.1. *Suppose that $\{\chi_n^\epsilon, \epsilon > 0, n < \infty\}$ and $\{\chi_n, n < \infty\}$ are uniformly integrable and $\{X_n\}$, $\{X_n^\epsilon\}$ are tight. Then $\{X^\epsilon(\cdot), \epsilon > 0\}$, $\{X^\epsilon(\epsilon q_\epsilon + \cdot), \epsilon > 0\}$, and $\{X^n(\cdot), n < \infty\}$ are tight and any weak limit has Lipschitz continuous paths w.p.1.*

Random measures. The treatment of the unbounded noise case (which is generally the situation in the problems of interest here) will be simplified and extended (over that in [22, 27]) by the use of random variables which are measure valued. The concept will be used in a rather simple way and all that we need to know will now be stated.

Let $\mathcal{P}(S)$ denote the set of probability measures over the Borel subsets of S . The Prohorov metric [10] will be used on this space. An important point is that under this metric $\mathcal{P}(S)$ is a CSMS, since S is [10, p. 101]. Convergence $P_n \rightarrow P$ in this topology is equivalent to weak convergence of $\{P_n, n < \infty\}$ to P [10, p. 108].

Now let $\{R_n, n < \infty\}$ be a sequence of *random variables* whose values are points in $\mathcal{P}(S)$. By definition, $\{R_n, n < \infty\}$ converges weakly (as a sequence of random variables) to the measure-valued random variable R if $EF(R_n) \rightarrow EF(R)$ for each bounded and continuous real-valued function $F(\cdot)$ on $\mathcal{P}(S)$. The function $\bar{R}_n$ defined by $\bar{R}_n = ER_n$ is a measure in $\mathcal{P}(S)$. We will need the following important fact

LEMMA 2.2 (see [24, pp. 14–15]). *The set $\{R_n, n < \infty\}$ has a weakly convergent subsequence if $\{\bar{R}_n, n < \infty\}$ is tight.*

This characterization in terms of the mean values is of great help, since the mean values $\{\bar{R}_n, n < \infty\}$ are much easier to deal with. Recall that the sequence $\{\bar{R}_n, n < \infty\}$ is tight if the associated sequence of random variables is tight. Let $f(\cdot)$ be a bounded, continuous, and real-valued function on S . The function defined by $F_f(P) = \int f(x)P(dx)$ is real valued, bounded, and continuous on $\mathcal{P}(S)$. Thus, if $R_n \Rightarrow R$ then $F_f(R_n) \rightarrow F_f(R)$ in distribution for each $f(\cdot)$. If the Skorohod representation is used, then we can say that w.p.1 for each such $f(\cdot)$

$$(2.3) \quad F_f(R_n) \rightarrow F_f(R).$$

3. The basic SA algorithms. The section contains several parts. Section 3.1 gives the main convergence theorem from which all others will be derived. Section 3.2 concerns the limit points of the ODEs which characterize the asymptotics of the SA. In many cases, the observation has a certain decomposition property which simplifies the verification of the assumptions, and this is exploited in §3.3. A simplified result for regenerative type processes, where we update at the end of the regeneration intervals, is in §3.4.

3.1. The canonical algorithm. Let $\epsilon > 0$. We will develop the basic ideas for the algorithm

$$(3.1) \quad \theta_{n+1}^\epsilon = \theta_n^\epsilon + \epsilon Y_n^\epsilon, \quad n \geq 0, \theta_n^\epsilon \in R^r,$$

where Y_n^ϵ is a sequence of R^r -valued random variables. The proofs of subsequent results for other SA forms will be more or less simple variations of the proof for (3.1). We next state the conditions which will be needed. The conditions seem to be nearly minimal and will be illustrated in the examples in §§7–9.

Let $\mathcal{B}_n^\epsilon$ be a sequence of nondecreasing sequence of sigma-algebras where $\mathcal{B}_n^\epsilon$ measures at least $\{\theta_0^\epsilon, Y_i^\epsilon, i < n\}$ and E_n^ϵ be the expectation conditioned on $\mathcal{B}_n^\epsilon$. Write $E_n^\epsilon Y_n^\epsilon = \bar{Y}_n^\epsilon$. Then the δY_n^ϵ defined by $Y_n^\epsilon = \bar{Y}_n^\epsilon + \delta Y_n^\epsilon$ are $\mathcal{B}_n^\epsilon$ -martingale differences. Generally, $\mathcal{B}_n^\epsilon$ will measure “all the information” which is used to get the $\{Y_i^\epsilon, i < n\}$. Suppose that

$$(3.2) \quad \{Y_n^\epsilon, n < \infty, \epsilon > 0\} \text{ is uniformly integrable.}$$

Suppose that there is a process $\{\xi_n^\epsilon, n < \infty\}$ which takes values in some CSMS and measurable functions $G_n^\epsilon(\cdot)$ such that we can write

$$(3.3) \quad \bar{Y}_n^\epsilon = G_n^\epsilon(\theta_n^\epsilon, \xi_n^\epsilon).$$

Assume that

$$(3.4) \quad \text{the set } \{\xi_n^\epsilon, \theta_n^\epsilon, \epsilon > 0, n < \infty\} \text{ is tight.}$$

Tightness of $\{\theta_n^\epsilon, \epsilon > 0, n < \infty\}$ holds if a projection algorithm is used (§5). Otherwise a stability argument might need to be used. Suppose that for each ϵ, θ, n there is a transition function $P_n^\epsilon(\cdot, \cdot | \theta)$ such that $P_n^\epsilon(\cdot, A | \cdot)$ is measurable for each Borel set A in the range space of ξ and

$$(3.5) \quad P\{\xi_{n+1}^\epsilon \in \cdot | \xi_i^\epsilon, \theta_i^\epsilon, i \leq n\} = P_n^\epsilon(\xi_n^\epsilon, \cdot | \theta_n^\epsilon).$$

By this Markov assumption, E_n^ϵ is the expectation conditioned on $(\theta_n^\epsilon, \xi_n^\epsilon)$. For each fixed θ , let there be a transition function $P(\xi, \cdot | \theta)$ such that

$$(3.6) \quad P_n^\epsilon(\xi, \cdot | \theta) \Rightarrow P(\xi, \cdot | \theta) \text{ as } n \rightarrow \infty, \epsilon \rightarrow 0,$$

where the limit is uniform on each compact (θ, ξ) set; i.e., for each bounded and continuous real-valued function $f(\cdot)$,

$$\int f(\tilde{\xi}) P_n^\epsilon(\xi, d\tilde{\xi} | \theta) \rightarrow \int f(\tilde{\xi}) P(\xi, d\tilde{\xi} | \theta)$$

uniformly on each compact (θ, ξ) set. Assume

$$(3.7) \quad P(\xi, \cdot | \theta) \text{ is weakly continuous in } (\theta, \xi).$$

For each fixed θ the transition function $P(\cdot, \cdot | \theta)$ determines a Markov chain and we let $\{\xi_n(\theta)\}$ denote the associated random variables. Let $\mu(\cdot | \theta)$ denote the invariant measures under the transition function $P(\xi, \cdot | \theta)$. Suppose that

$$(3.8) \quad \{\mu(\cdot | \theta), \theta \in \Theta\} \text{ is tight for each compact } \Theta.$$

Henceforth, let q_ϵ be a sequence of integers such that

$$(3.8') \quad \begin{array}{ll} \text{either} & q_\epsilon \equiv 0 \\ \text{or} & \epsilon q_\epsilon \rightarrow \infty. \end{array}$$

Suppose that there is a continuous function $G(\cdot)$ such that for each $\delta > 0$

$$(3.9) \quad \lim_{\epsilon} \limsup_n P\{|G_n^\epsilon(\theta_n^\epsilon, \xi_n^\epsilon) - G(\theta_n^\epsilon, \xi_n^\epsilon)| \geq \delta\} = 0$$

and that for each compact θ -set Θ there is $K_0(\Theta) < \infty$ such that for all stationary processes $\{\xi_n(\theta)\}$

$$(3.10) \quad \sup_{\theta \in \Theta} E|G(\theta, \xi_j(\theta))| < K_0(\Theta).$$

Finally, we assume either (3.11a) or (3.11b):

$$(3.11a) \quad \text{For each } \theta, \mu(\cdot|\theta) \text{ is unique.}$$

There is a continuous $g(\cdot)$ such that for each θ and initial condition $\xi_0(\theta)$

$$(3.11b) \quad \lim_N \frac{1}{N} \sum_{n=0}^{N-1} EG(\theta, \xi_n(\theta)) = g(\theta).$$

Under (3.11a), define

$$g(\theta) = \int G(\theta, \xi) \mu(d\xi|\theta).$$

Define the continuous parameter interpolation $\theta^\epsilon(\cdot)$ by $\theta^\epsilon(t) = \theta_n^\epsilon$ for $t \in [n\epsilon, n\epsilon + \epsilon)$, $n \geq 0$. For $t < 0$, set $\theta^\epsilon(t) = \theta_0^\epsilon$.

THEOREM 3.1. *Assume the conditions (3.2)–(3.11). Each subsequence of $\{\theta^\epsilon(q_\epsilon\epsilon + \cdot), \epsilon > 0\}$ has a further subsequence which converges weakly to a bounded solution $\theta(\cdot)$ of*

$$(3.12) \quad \dot{\theta} = g(\theta)$$

on $[0, \infty)$ if $q_\epsilon = 0$ and on $(-\infty, \infty)$ if $\epsilon q_\epsilon \rightarrow \infty$. Also, $g(\cdot)$ is a continuous function of θ .

Remark. We note that in current applications it is often the case that the P_n^ϵ and the G_n^ϵ do not depend on either ϵ or n . See the examples in §§6–8. A way of avoiding the Markovianization is described in Appendix 1. Condition (3.11b) is often much easier to check than is uniqueness of the invariant measure. In typical examples where one uses some sort of weak sense derivative or an IPA-type estimator, it is equivalent to the asymptotic consistency of the estimator under fixed θ , as will be seen in the examples in §§6–8. This is a minimal condition. The ability to use such a condition is basically a consequence of the “martingale problem” formulation used in the proof. It is exploited in the use of conditional expectations in the expressions from (3.17) on.

The basic idea in the proof is to first replace the Y_n^ϵ by its conditional expectation, given the past. Then use a piecewise constant approximation to the state process, and finally exploit this last approximation via an ergodic condition. The type of continuity

and uniform integrability conditions required seem rather weak and have their roots in the basic references [22, 27].

Remark. If Y_n^ϵ can be represented as $g(\theta_n^\epsilon)$ plus a “martingale difference” plus a term which goes to zero in mean¹ as $\epsilon \rightarrow 0$ and/or $n \rightarrow \infty$, then the proof becomes nearly trivial since no averaging needs to be done. The difficulties arise when the conditional expectation (given past data) of Y_n^ϵ depends on the past, and this holds true in many important cases. The basic structure and motivation of the proof are analogous to those of [22, 27], but many of the details are different. Here there is a smoother development of the unbounded noise case under weaker conditions. The proof also provides a simpler way of characterizing the limit points (see Theorem 3.2) and dealing with the other extensions. In order to simplify the notation, we use $q_\epsilon = 0$ in the proof. The details are exactly the same for the general case.

Proof. Part 1. A continuity result. Until the last part of the proof, assume (3.11a). Let $f(\cdot)$ be bounded, continuous, and real valued. Given $\theta_0 \in R^r$, let θ_n be a deterministic sequence tending to θ_0 . We have

$$\int f(\xi)\mu(d\xi|\theta_n) = \int \left[\int f(\tilde{\xi})P(\xi, d\tilde{\xi}|\theta_n) \right] \mu(d\xi|\theta_n).$$

Now as $n \rightarrow \infty$ $P(\xi, \cdot|\theta_n)$ converges weakly to $P(\xi, \cdot|\theta_0)$ uniformly on each compact (θ, ξ) set by (3.7). Using (3.8), extract a weakly convergent subsequence of $\{\mu(\cdot|\theta_n), n < \infty\}$ and denote the limit by $\tilde{\mu}(\cdot)$. Then

$$\int f(\xi)\tilde{\mu}(d\xi) = \int \left[\int f(\tilde{\xi})P(\xi, d\tilde{\xi}|\theta_0) \right] \tilde{\mu}(d\xi),$$

which implies, via uniqueness, that $\tilde{\mu}(\cdot) = \mu(\cdot|\theta_0)$. This argument yields the continuity of $\int f(\xi)\mu(d\xi|\theta)$.

Part 2. A martingale problem representation. By (3.1) and (3.3)

$$(3.13) \quad \theta^\epsilon(t) = \theta_0^\epsilon + \epsilon \sum_{i=0}^{t/\epsilon-1} G_i^\epsilon(\theta_i^\epsilon, \xi_i^\epsilon) + \epsilon \sum_{i=0}^{t/\epsilon-1} \delta Y_i^\epsilon.$$

First, we show that the martingale term (the one on the right) goes to zero as $\epsilon \rightarrow 0$. This would be easy if the uniform integrability in (3.2) were replaced by square integrability, since then the martingale would be square integrable and its variance at t would be bounded by $\epsilon^2(t/\epsilon) \sup_{\epsilon, n} \text{var}(\delta Y_n^\epsilon) = O(\epsilon t)$. Hence the term would have the zero process as a weak limit. We get the same result by a truncation argument, as follows. For large positive B , let $I_{n,B}^\epsilon$ be the indicator function of the event that Y_n^ϵ does not exceed B in absolute magnitude. Then use $Y_n^\epsilon I_{n,B}^\epsilon$ in lieu of Y_n^ϵ , as follows. Define $\delta Y_{n,B}^\epsilon$ and $\beta_{n,B}^\epsilon$ by

$$Y_n^\epsilon I_{n,B}^\epsilon = E_n Y_n^\epsilon I_{n,B}^\epsilon + \delta Y_{n,B}^\epsilon, \quad Y_n^\epsilon = Y_n^\epsilon I_{n,B}^\epsilon + \beta_{n,B}^\epsilon.$$

We have $\sup_{\epsilon, n} E|\beta_{n,B}^\epsilon| \rightarrow 0$ as $B \rightarrow \infty$ by the uniform integrability. Since $\{\delta Y_{n,B}^\epsilon\}$ are bounded, for each $B < \infty$ the martingale term $\epsilon \sum_{i=0}^{t/\epsilon-1} \delta Y_{i,B}^\epsilon$ contributes nothing to the limit by the “square integrability” theory. Now the uniform integrability (3.2) yields

$$\limsup_{B, n, \epsilon} E|E_n Y_n^\epsilon I_{n,B}^\epsilon - G_n^\epsilon(\theta_n^\epsilon, \xi_n^\epsilon)| = 0.$$

¹ If, for a sequence Z_n , $E|Z_n| \rightarrow 0$, we say that it converges in mean to zero.

These results imply that (3.13) can be written as

$$(3.14) \quad \theta^\epsilon(t) = \theta^\epsilon(0) + \epsilon \sum_{j=0}^{t/\epsilon-1} G_j^\epsilon(\theta_j^\epsilon, \xi_j^\epsilon) + \rho^\epsilon(t),$$

where $|\rho^\epsilon(t)| \rightarrow 0$ in the mean uniformly on each bounded t -interval. Now the uniform integrability (3.2) and the form (3.1) imply that $\{\theta^\epsilon(\cdot), \epsilon > 0\}$ is tight and that any weak limit has Lipschitz-continuous paths w.p.1 (see Lemma 2.1).

The conditions seem to be weakest if we work with a “martingale problem” formulation, and we proceed to do so. Now, with a slight abuse of notation, let ϵ index a weakly convergent subsequence of $\{\theta^\epsilon(\cdot), \epsilon > 0\}$ with limit process denoted by $\theta(\cdot)$. Let t, τ be arbitrary positive numbers; q be an integer; $s_i, i \leq q$, be nonnegative numbers no larger than t ; and $h(\cdot)$ be a bounded, continuous, and real-valued function of its arguments. As is common in weak convergence-type arguments, we will show that

$$(3.15) \quad Eh(\theta(s_i), i \leq q) \left[\theta(t + \tau) - \theta(t) - \int_t^{t+\tau} g(\theta(u)) du \right] = 0.$$

By the arbitrariness of the $h(\cdot), q, t, \tau, s_i$, (3.15) implies that $\theta(t) - \theta(0) - \int_0^t g(\theta(u)) du$ is a martingale (with respect to the filtration which it generates). Since $E|\rho^\epsilon(\cdot)| \rightarrow 0$ and $\{G_n^\epsilon(\theta_n^\epsilon, \xi_n^\epsilon), \epsilon > 0, n < \infty\}$ is uniformly integrable, the form (3.14) implies that the martingale has zero quadratic variation; hence it is constant. Since it takes the value zero at $t = 0$, it is identically zero w.p.1. Thus, the theorem will be proved once (3.15) is proved.

Part 3. Approximating the $G_j^\epsilon(\cdot)$. By the properties of $\rho^\epsilon(\cdot)$, we can write

$$(3.16) \quad Eh(\theta^\epsilon(s_i), i \leq q) \left[\theta^\epsilon(t + \tau) - \theta^\epsilon(t) - \epsilon \sum_{j=t/\epsilon}^{(t+\tau)/\epsilon-1} G_j^\epsilon(\theta_j^\epsilon, \xi_j^\epsilon) \right] \rightarrow 0$$

as $\epsilon \rightarrow 0$. We proceed to rearrange the terms in (3.16) so that efficient averaging methods can be used. Let $n_\epsilon \rightarrow \infty$ be a sequence of integers such that $\delta_\epsilon = \epsilon n_\epsilon \rightarrow 0$ and τ is an integral multiple of δ_ϵ . Without loss of generality and for notational simplicity, suppose that t is also an integral multiple of δ_ϵ . By collecting terms in groups of size n_ϵ and using the freedom that we have with taking the conditional expectations given “past data” inside the brackets in (3.16), we can write the left side of (3.16) as

$$(3.17) \quad Eh(\theta^\epsilon(s_i), i \leq q) \left\{ \theta^\epsilon(t + \tau) - \theta^\epsilon(t) - \sum_{l: l\delta_\epsilon=t}^{t+\tau-\delta_\epsilon} \delta_\epsilon \left[\frac{1}{n_\epsilon} \sum_{j=ln_\epsilon}^{ln_\epsilon+n_\epsilon-1} E_{ln_\epsilon}^\epsilon G_j^\epsilon(\theta_j^\epsilon, \xi_j^\epsilon) \right] \right\}.$$

For a real-valued function $f(\cdot)$ and $\infty > B > 0$, define $f_B(\cdot)$ by $f_B(x) = \min[f(x), B]$ for $f(x) \geq 0$ and by $f_B(x) = \max[f(x), -B]$ otherwise.

By the uniform integrability (3.2), given any $\rho > 0$ there is $B < \infty$ such that we can use $G_{n_\epsilon, B}^\epsilon(\cdot)$ while changing the expectations of the absolute values of the summands in the brackets of (3.17) by at most ρ . Continuing in the bracketed term, first replace $G_j^\epsilon(\theta_j^\epsilon, \xi_j^\epsilon)$ with $G_{j, B}^\epsilon(\theta_j^\epsilon, \xi_j^\epsilon)$ plus a small error term. Then use (3.9) (which also holds for the B -truncated functions) to replace $G_{j, B}^\epsilon(\theta_j^\epsilon, \xi_j^\epsilon)$ with $G_B(\theta_j^\epsilon, \xi_j^\epsilon)$ plus

a small error. Finally, use the (uniform in compact (θ, ξ) sets) continuity of $G_B(\cdot, \xi)$, (3.4) and the fact that (in probability, uniformly in l)

$$(3.18) \quad \sup_{j \leq n_\epsilon} |\theta_{ln_\epsilon+j}^\epsilon - \theta_{ln_\epsilon}^\epsilon| \rightarrow 0 \quad \text{as } \epsilon \rightarrow 0$$

to justify replacing θ_j^ϵ by $\theta_{ln_\epsilon}^\epsilon$ (plus a small error term), yielding that (3.17) equals (3.19)

$$Eh(\theta^\epsilon(s_i), i \leq q) \left\{ \theta^\epsilon(t + \tau) - \theta^\epsilon(t) - \sum_{l: l\delta_\epsilon = t}^{t+\tau-\delta_\epsilon} \delta_\epsilon \left[\frac{1}{n_\epsilon} \sum_{j=ln_\epsilon}^{ln_\epsilon+n_\epsilon-1} E_{ln_\epsilon}^\epsilon G_B(\theta_{ln_\epsilon}^\epsilon, \xi_j^\epsilon) \right] \right\}$$

modulo an error ρ' which can be made as small as desired in mean value by choosing B large enough and then ϵ small enough. The sum in (3.19) can be written as $\int_t^{t+\tau} \tilde{G}_B^\epsilon(s) ds$, with the obvious definition of $\tilde{G}_B^\epsilon(\cdot)$ as the process which is constant on intervals $[l\delta_\epsilon, (l+1)\delta_\epsilon)$, as defined² by the bracketed term.

The weak convergence arguments in the next parts will show that

$$(3.20) \quad Eh(\theta^\epsilon(s_i), i \leq q) \tilde{G}_B^\epsilon(s) \rightarrow Eh(\theta(s_i), i \leq q) g_B(\theta(s)),$$

where $g_B(\theta) = \int G_B(\theta, \xi) \mu(d\xi|\theta)$. This will imply that the outer sum in (3.19) can be replaced by $\int_t^{t+\tau} g_B(\theta(s)) ds$ in the limit as $\epsilon \rightarrow 0$. These results will yield that

$$(3.21) \quad \dot{\theta} = g_B(\theta) + \rho_B,$$

where $E \int_0^t |\rho_B(s)| ds \rightarrow 0$ as $B \rightarrow \infty$. The proof under (3.11a) will then be completed in part 5 by showing that we can let $B = \infty$. Part 6 will deal with (3.11b).

Part 4. Averaging out the ξ_j^ϵ . To complete our program, we need to average out the ξ_j^ϵ terms in (3.19). Define the measure-valued random variable (an average of conditional probabilities)

$$(3.22) \quad R(l, \epsilon, \cdot) = \frac{1}{n_\epsilon} \sum_{j=ln_\epsilon}^{ln_\epsilon+n_\epsilon-1} P\{\xi_j^\epsilon \in \cdot | \theta_{ln_\epsilon}^\epsilon, \xi_{ln_\epsilon}^\epsilon\},$$

and recall that the $E_{ln_\epsilon}^\epsilon$ is the expectation conditioned on $(\theta_{ln_\epsilon}^\epsilon, \xi_{ln_\epsilon}^\epsilon)$ by the Markov assumption. The inner square bracketed term in (3.19) can now be written

$$(3.23) \quad \int R(l, \epsilon, d\xi) G_B(\theta_{ln_\epsilon}^\epsilon, \xi).$$

The set of measure-valued random variables $\{R(l, \epsilon, \cdot), l < \infty, \epsilon > 0\}$ is tight, since the mean values are just

$$\bar{R}(l, \epsilon, \cdot) = \frac{1}{n_\epsilon} \sum_{j=ln_\epsilon}^{ln_\epsilon+n_\epsilon-1} P\{\xi_j^\epsilon \in \cdot\}$$

² We note here that the major problem in averaging the inner square bracket is in showing that the ξ_j^ϵ in the inner sum in (3.19) can be replaced by $\xi_j(\theta_{ln_\epsilon}^\epsilon)$, all of which have the same value of θ as an argument, so that some sort of ergodic theorem or averaging principle can be used to get the ultimate averaged limit. Recall that $\xi_i(\theta)$ is the Markov chain with fixed parameter θ . This idea is basic to all of the averaging methods. The continuity (3.7) of the transition function is the basic property that is used. Various alternatives will appear in the appendices.

and the tightness of $\{\bar{R}(l, \epsilon, \cdot), l \geq 0, \epsilon > 0\}$ is just the tightness of $\{\xi_n^\epsilon, \epsilon > 0, n < \infty\}$, and this latter sequence is tight by assumption (3.4) (see Lemma 2.2). We will characterize the limits of weakly convergent subsequences of $\{R(l, \epsilon, \cdot), l < \infty, \epsilon > 0\}$ as measures whose values are (w.p.1) just the invariant measure $\mu(\cdot|\theta)$ with appropriate values of θ . Now we follow the ideas in the development in [22, p. 110], except for the use of the random measures in place of their pointwise values. (We note that the use of random measures here greatly simplifies the treatment of the unbounded noise case over that in the references.)

Fix $s > 0$, and let l_ϵ be such that $s \in [l_\epsilon \delta_\epsilon, l_\epsilon \delta_\epsilon + \delta_\epsilon)$ for all ϵ . Let $f(\cdot)$ be a bounded and continuous real-valued function of ξ . Using (3.8), extract a weakly convergent subsequence of $\{R(l_\epsilon, \epsilon, \cdot), \theta^\epsilon(\cdot), \epsilon > 0\}$, and index it by $\epsilon(p), p \rightarrow \infty$. The proof will show that this further subsequence is irrelevant, due to the uniqueness of the $\mu(\cdot|\theta)$, and that we can let $\epsilon(p) = \epsilon$. We also suppose that the Skorohod representation is used so that the convergences are w.p.1 in the appropriate topologies. Denote the limit by $(\bar{R}(\cdot), \theta(\cdot))$.³ Define $m_{\epsilon(p)} = l_{\epsilon(p)} n_{\epsilon(p)}$. Note that $m_{\epsilon(p)} \rightarrow \infty$ as $p \rightarrow \infty$. We can write

$$(3.24) \quad \begin{aligned} \int \bar{R}(d\xi) f(\xi) &= \lim_{p \rightarrow \infty} \int R(l_{\epsilon(p)}, \epsilon(p), d\xi) f(\xi) \\ &= \lim_{p \rightarrow \infty} \frac{1}{n_{\epsilon(p)}} \sum_{j=m_{\epsilon(p)}}^{m_{\epsilon(p)}+n_{\epsilon(p)}-1} \int P\{\xi_j^{\epsilon(p)} \in d\xi | \theta_{m_{\epsilon(p)}}^{\epsilon(p)}, \xi_{m_{\epsilon(p)}}^{\epsilon(p)}\} f(\xi). \end{aligned}$$

The first equality follows from the definition of the weak limit. The second follows from definition of $R(l_\epsilon, \epsilon, \cdot)$. Continuing, we use the one-step transition function $P_j^{\epsilon(p)}(\cdot)$ to rewrite the right side of (3.24) as (minus the first term of the sum)

$$(3.25) \quad \lim_p \frac{1}{n_{\epsilon(p)}} \sum_{j=m_{\epsilon(p)}+1}^{m_{\epsilon(p)}+n_{\epsilon(p)}-1} \int \int P\{\theta_{j-1}^{\epsilon(p)} \in d\tilde{\theta}, \xi_{j-1}^{\epsilon(p)} \in d\tilde{\xi} | \theta_{m_{\epsilon(p)}}^{\epsilon(p)}, \xi_{m_{\epsilon(p)}}^{\epsilon(p)}\} P_{j-1}^{\epsilon(p)}(\tilde{\xi}, d\tilde{\xi} | \tilde{\theta}) f(\xi).$$

Condition (3.6) yields

$$\lim_{p, n} \int P_n^{\epsilon(p)}(\tilde{\xi}, d\tilde{\xi} | \tilde{\theta}) f(\xi) = \int P(\tilde{\xi}, d\tilde{\xi} | \tilde{\theta}) f(\xi) \equiv \bar{f}(\tilde{\theta}, \tilde{\xi}),$$

and the limit is uniform on each compact $(\tilde{\theta}, \tilde{\xi})$ set. By (3.7), the right-hand side is continuous. By using these facts and the fact that the limit $\theta(\cdot)$ is continuous, we can concentrate the measure of $\tilde{\theta}$ in (3.25) at $\theta_{m_{\epsilon(p)}}^{\epsilon(p)}$ without affecting the limit. With this replacement, (3.25) can be written as

$$\lim_p \int R(l_{\epsilon(p)}, \epsilon(p), d\xi) \bar{f}(\theta_{m_{\epsilon(p)}}^{\epsilon(p)}, \xi),$$

which by the use of the weak convergence of $\{R^{\epsilon(p)}(l_{\epsilon(p)}, \epsilon(p), \cdot), \theta_{m_{\epsilon(p)}}^{\epsilon(p)}\}$ to $(\bar{R}(\cdot), \theta(s))$ as $p \rightarrow \infty$ equals

$$(3.26) \quad \int \bar{R}(d\xi) \bar{f}(\theta(s), \xi) = \int \int \bar{R}(d\tilde{\xi}) P(\tilde{\xi}, d\tilde{\xi} | \theta(s)) f(\xi).$$

³ Strictly speaking, when taking limits of $\{\bar{R}(l_{\epsilon(p)}, \epsilon(p), \cdot), \theta^{\epsilon(p)}(\cdot), p < \infty\}$ and using Skorohod representation, the probability space might be different from what was used when we got the original weakly convergent subsequence with limit $\theta(\cdot)$ in part 2. But since $\{\theta^{\epsilon(p)}(\cdot), p < \infty\}$ is a subsequence of $\{\theta^\epsilon(\cdot), \epsilon > 0\}$ and all that matters are the distributions of the resulting limits anyway, we write $\theta(\cdot)$ for the limit as $p \rightarrow \infty$ for notational simplicity and without loss of generality.

Let ω denote the canonical probability space variable. Equating the right side of (3.26) with the left-hand side of (3.24) yields that (w.p.1) each sample value $\tilde{R}(\cdot, \omega)$ must be an invariant measure for the transition probability $P(\xi, \cdot | \theta(s, \omega))$. By the uniqueness of the invariant measure, $\tilde{R}(\cdot) = \mu(\cdot | \theta(s))$ w.p.1, and the subsequence of $\{R(l_\epsilon, \epsilon, \cdot), \epsilon > 0\}$ which is used is irrelevant. This implies that the limit in (3.23) is $g_B(\theta(s))$ w.p.1, which yields (3.20).

Part 5. Replacing $g_B(\cdot)$ by $g(\cdot)$. The results of the previous parts imply that we can replace the square bracketed term of (3.20) by (3.19) by $g_B(\theta(s))$ and that (3.20), (3.21) hold. We need only show that we can let $B \rightarrow \infty$ in (3.21). We can let $B \rightarrow \infty$ and replace $g_B(\cdot)$ by $g(\cdot)$ if $G(\theta, \cdot)$ is $\mu(\cdot | \theta)$ integrable for each θ , the integral is bounded on each compact θ set, and

$$(3.27) \quad \int G_B(\theta, \xi) \mu(d\xi | \theta) \rightarrow \int G(\theta, \xi) \mu(d\xi | \theta)$$

uniformly on each compact θ set. But this follows from (3.10) and the monotone convergence theorem.

Part 6. Using (3.11b). If we drop the uniqueness condition, then the subsequence $\epsilon(p)$ might be important. However, note that the above proof established that, whether or not there is uniqueness,

$$\dot{\theta}(s) = \int G(\theta(s), \xi) \mu(d\xi | \theta(s))$$

for some invariant measure, which might depend on (ω, s) . But (3.11b) implies that the right-hand side equals $g(\theta(s))$ for the function $g(\cdot)$ defined there. $\square$

3.2. Limit points and nonunique invariant measures: Limit points of (3.12). We use some elementary facts from the theory of differential equations. Given an ODE $\dot{x} = f(x)$, $x \in R^n$, with continuous $f(\cdot)$ and a bounded solution $x(\cdot)$ on $[0, \infty)$, let L denote the set of limit points of the path $x(\cdot)$. Define an *invariant set* M for the ODE as follows. For each $y \in M$, there is a solution $y(\cdot)$ to the ODE on $(-\infty, \infty)$ such that $y(t) \in M$ for all t and $y(0) = y$. Then [15] L is a compact invariant set. We can now state the following result.

THEOREM 3.2. *Assume the conditions of Theorem 3.1. The limit points of (3.12) are contained in the largest bounded invariant set M of $\dot{x} = g(x)$. Now let $\epsilon q_\epsilon \rightarrow \infty$ and $\theta^\epsilon(\epsilon q_\epsilon + \cdot) \Rightarrow \theta(\cdot)$ as $\epsilon \rightarrow 0$. Then, w.p.1 for each $t \in (-\infty, \infty)$, $\theta(t) \in M$.*

Remark. The last assertion holds since the solution is defined on the doubly infinite time interval. If the ODE has a unique stationary point $\hat{\theta}$, then the last assertion implies that $\theta(t) = \hat{\theta}$ for all t . Appropriate perturbation schemes will guarantee that the iterate won't get stuck at a maximum or at a saddle point. Reference [1] shows that the set of limit points are confined to the set of chain recurrent points, which might be smaller than the largest bounded invariant set, but the conditions are stronger.

Nonunique invariant measure. Suppose that $\mu(\cdot | \theta)$ is not unique and (3.11b) cannot be verified. We might still be able to get a useful result. Let $V(\theta)$ denote the (convex) set of invariant measures under θ . The proof of Theorem 3.1 can be easily modified to get the following theorem.

THEOREM 3.3. *Assume the conditions of Theorem 3.1 except for (3.11). Then w.p.1 and for almost all t , the theorem holds with (3.12) replaced by*

$$(3.28) \quad \dot{\theta}(t) \in \left\{ \int G(\theta(t), \xi) \mu(d\xi | \theta(t)), \mu(\cdot | \theta(t)) \in V(\theta(t)) \right\}.$$

3.3. “Atomic” increments. Return to the basic algorithm (3.1). In many applications the Y -variables have an additivity property which can simplify the verification of the conditions and which we now explain and exploit. Define $q_0^\epsilon = 0$ and suppose that we update at the increasing random times $q_n^\epsilon, n = 1, \dots$. Let $k^\epsilon(n)$ be the last time of updating before and including time n . By additivity, we mean that the observations can be divided up such that the algorithm can be written

$$(3.29) \quad \theta_{n+1}^\epsilon = \theta_n^\epsilon + \epsilon \sum_{i=q_n^\epsilon}^{q_{n+1}^\epsilon-1} Y_i^\epsilon,$$

where the Y_n^ϵ obey the conditions of Theorem 3.1. At each instant i in the interval $[q_n^\epsilon, q_{n+1}^\epsilon)$ the value θ_n^ϵ is used to get Y_i^ϵ . From the point of view of the convergence theory, one can just as well update in “real time” and use the modified algorithm

$$(3.30) \quad \theta_{n+1}^\epsilon = \theta_n^\epsilon + \epsilon Y_n^\epsilon,$$

where $\theta_{k^\epsilon(n)}^\epsilon$ is used to get Y_n^ϵ . Suppose that $q_{n+1}^\epsilon - q_n^\epsilon$ is bounded by some constant independently of ϵ, n . Then the conditions of Theorem 3.1 guarantee that its conclusions hold for the SA (3.30) and similarly for its extensions. If the intervals $q_{n+1}^\epsilon - q_n^\epsilon$ are unbounded, then in order for the two time scales of (3.29) and (3.30) to be compatible, we need in addition that the conditions of Theorem 3.1 hold for the original algorithm, in particular that $[\theta_{n+1}^\epsilon - \theta_n^\epsilon]/\epsilon$ be uniformly (in n, ϵ) integrable.

The advantage of this “atomic” decomposition is that it makes it easier to verify the conditions on the Markov chain ξ_n^ϵ for (3.30) than for (3.39), since the transitions are viewed “more locally.”

3.4. Updating at regeneration times. Suppose that the problem has a structure that allows $g(\theta)$ to be estimated regeneratively. When $g(\theta)$ is the derivative of $C(\theta)$, a continuously differentiable, “stationary cost” function of a regenerative process, an excellent treatment of the regenerative estimation of the derivative is in [14]. We might wish to use SA to minimize $C(\theta)$. In particular, suppose that θ_n^ϵ will be updated at the end of each new k regeneration intervals and that there are Y_n^ϵ which are “nearly” unbiased estimators of $g(\theta_n^\epsilon)$ and which depend on data in regeneration intervals $[nk + 1, nk + k]$ only. Let (3.1) be used. Define

$$G_n^\epsilon(\theta) = E[Y_n^\epsilon | \theta_n^\epsilon = \theta, \theta_i^\epsilon, i < n] = E[Y_n^\epsilon | \theta_n^\epsilon = \theta].$$

Assume that $\{\theta_n^\epsilon, \epsilon > 0, n < \infty\}$ is tight, $\{Y_n^\epsilon, \epsilon > 0, n < \infty\}$ is uniformly integrable, and, for each $\delta > 0$,

$$(3.31) \quad \lim_{\epsilon} \limsup_n P\{|G_n^\epsilon(\theta_n^\epsilon) - g(\theta_n^\epsilon)| \geq \delta\} = 0.$$

Then a simpler proof than that of Theorem 3.1 says that the conclusions of Theorem 3.1 hold (and similarly for the other theorems). The proof is simpler since there is no need to introduce ξ_n^ϵ or the averaging measures. We note that sometimes the minimization of an average cost per unit time can be reduced to an SA iteration where we update at the end of each k intervals for some integer k . See, for example, §2.2 of [32]. The results of Theorem 3.1, as expanded in Appendices 3 and 4, show that it is not necessary to use regeneration intervals as the basis for updating. Indeed, updating only at the ends of these intervals might be a poor idea in practice in general, despite the fact that the proofs are simplified. For network problems, a regeneration

interval-based approach would be a handicap, since the intervals would generally be very long. The situation is even worse if the processing is distributed (Appendix 4). The same point was made by [35].

4. Time-varying gains $\epsilon_n \rightarrow 0$. Suppose that the positive real numbers ϵ_n go to zero such that

$$(4.1a) \quad \sum_{n=0}^{\infty} \epsilon_n = \infty$$

and

$$(4.1b) \quad \text{either } \sum_n |\epsilon_{n+1} - \epsilon_n| < \infty \text{ or } \epsilon_n / \epsilon_{n+1} \rightarrow 1.$$

These ϵ_j could actually be random if they are nonanticipative and satisfy (4.1). The SA algorithm is

$$(4.2) \quad \theta_{n+1} = \theta_n + \epsilon_n Y_n.$$

Let $\mathcal{B}_n$ be a sequence of nondecreasing sigma-algebras measuring at least $\{\theta_0, Y_i, i < n\}$. Write E_n for the conditional expectation given $\mathcal{B}_n$. The δY_n defined by $Y_n = E_n Y_n + \delta Y_n$ are $\mathcal{B}_n$ -martingale differences. As for Theorem 3.1, suppose that there is a process $\{\xi_n, n < \infty\}$ taking values in a CSMS and functions $G_n(\cdot)$ such that $E_n Y_n = G_n(\theta_n, \xi_n)$. Define $t_n = \sum_{i=0}^{n-1} \epsilon_i$. Following the definition (2.2) for $t \geq 0$, define $\theta^0(t) = \theta_n$ on $[t_n, t_{n+1})$ and set $\theta^0(t) = \theta_0$ on $(-\infty, 0]$. Define $\theta^n(t) = \theta^0(t_n + t)$. Thus $\theta^n(0) = \theta_n$.

Theorems 3.1–3.3 readily lead to the following theorem.

THEOREM 4.1. *Assume the conditions of Theorem 3.1 with (4.1), (4.2), and the following replacements. Equations (3.2)–(3.6) and (3.10) hold with the superscript ϵ dropped. Use $\limsup_n$ in (3.9). Then $\{\theta^n(\cdot), n < \infty\}$ is tight and the limit of any weakly convergent subsequence satisfies (3.12) on $(-\infty, \infty)$ w.p.1. Also, w.p.1., for all t , $\theta(t) \in M$, the largest bounded invariant set of (3.12). If (3.11) is dropped, the conclusions of Theorem 3.3 still hold. The obvious analogue of the results for the “atomic” increments formulation also hold.*

Remarks on the proof. Again, the general structure is similar to that used in [28] but with differing details. The proof is essentially the same as those of Theorems 3.1 and 3.3. The only difference concerns the way the terms are grouped, i.e., the analogy to the arrangement in (3.17). For simplicity, let $t \geq 0, \tau > 0$. Define $m_n(t) = \max\{j \geq n : t_j - t_n \leq t\}$. The following expression replaces (3.16):

$$(4.3) \quad Eh(\theta^n(s_i), i \leq q) \left[\theta^n(t + \tau) - \theta^n(t) - \sum_{j=m_n(t)}^{m_n(t+\tau)-1} \epsilon_j (G_j(\theta_j, \xi_j) + \delta Y_j) \right] = 0.$$

The martingale term

$$\sum_{j=m_n(t)}^{m_n(t+\tau)-1} \epsilon_j \delta Y_j$$

goes to zero as $n \rightarrow \infty$ by a bounding argument of the type used in Theorem 3.1 and the fact that $\epsilon_n \rightarrow 0$. Note in particular that no condition of the form $\sum \epsilon_j^{1+\delta} < \infty$

for $\delta > 0$ is needed. It is only required that $\epsilon_n \rightarrow 0$. Indeed, if the $\{Y_n, n < \infty\}$ were uniformly square integrable, then the variance of the martingale term would be $O(1) \sum_{m_n(t)}^{m_n(t+\tau)-1} \epsilon_j^2$ and this goes to zero if $\epsilon_j \rightarrow 0$, since $\sum_{j=m_n(t)}^{m_n(t+\tau)-1} \epsilon_j \approx \tau$. In general, one uses uniform integrability to the same end as in part 2 of the proof of Theorem 3.1.

We now comment briefly on the appropriate grouping of the terms. The δ_ϵ used for the grouping in (3.17) is replaced with a sequence of positive numbers $\delta_n \rightarrow 0$ which satisfy $\lim_n \sup\{\epsilon_j/\delta_n : j \geq n\} = 0$. Define $m(n, 0) = n$. For each n , define an increasing sequence of integers $m(n, l), l = 1, \dots$, by

$$m(n, l) = \min \left\{ j : \sum_{i=n}^{j-1} \epsilon_i \geq l\delta_n \right\}.$$

Thus

$$\sum_{j=m(n,l)}^{m(n,l+1)-1} \epsilon_j \approx \delta_n.$$

For each n , we will arrange the terms in groups of successive sizes $m(n, l+1) - m(n, l)$ as follows. Suppose, for notational simplicity, that both t and $t+\tau$ are integral multiples of δ_n . The changes for the general case should be obvious. Now, analogously to what was done in part 3 of the proof of Theorem 3.1, replace the sum of the $\epsilon_j G_j$ in (4.3) with

$$(4.4) \quad \sum_{l:l\delta_n=t}^{(t+\tau)-\delta_n} \delta_n \left[\frac{1}{\delta_n} E_{m(n,l)} \sum_{j=m(n,l)}^{m(n,l+1)-1} \epsilon_j G_j(\theta_j, \xi_j) \right].$$

Then argue that the $G_j(\theta_j, \xi_j)$ in the sum on the right can be replaced by $G_B(\theta_{m(n,l)}, \xi_j)$ plus an arbitrarily small (in the mean) error for large B, n . Define the measure-valued random variables $R(l, n, \cdot)$ by

$$(4.5) \quad R(l, n, \cdot) = \frac{1}{\delta_n} \sum_{j=m(n,l)}^{m(n,l+1)-1} \epsilon_j P\{\xi_j \in \cdot | \theta_{m(n,l)}, \xi_{m(n,l)}\}.$$

Thus, as in part 4 of the proof of Theorem 3.1, we approximate the bracketed term in (4.4) by

$$\int G_B(\theta_{m(n,l)}, \xi) R(l, n, d\xi).$$

Next consider the analogue of the factorization taking the sum on the right side of (3.24) to that in (3.25). The analogue of (3.25) (before extracting the convergent subsequence and without the limit) is

$$\frac{1}{\delta_n} \sum_{j=m(n,l)+1}^{m(n,l+1)-1} \int \int \epsilon_j P\{\theta_{j-1} \in d\tilde{\theta}, \xi_{j-1} \in d\tilde{\xi} | \theta_{m(n,l)}, \xi_{m(n,l)}\} P_{j-1}(\tilde{\xi}, d\tilde{\xi} | \tilde{\theta}) f(\xi).$$

Analogously to part 4 of the proof of Theorem 3.1, we can fix $\tilde{\theta}$ at $\theta_{m(n,l)}$ and replace $P_j(\cdot)$ with $P(\cdot)$ to get the representation

$$\int R(l, n, d\tilde{\xi}) \int f(\xi) P(\tilde{\xi}, d\xi | \theta_{m(n,l)}),$$

where ϵ_{j-1} replaces ϵ_j in the definition of $R(\cdot)$. By either option in (4.1b), this replacement does not affect the limit.

5. A constrained algorithm. Let H be a closed set in R^r . Let $\Pi_H(x)$ denote the closest point in H to x . The following development is a slight extension of [22, pp. 111–114]. Define the projected form of (3.1) as

$$(5.1) \quad \theta_{n+1}^\epsilon = \Pi_H(\theta_n^\epsilon + \epsilon Y_n^\epsilon).$$

Rewrite (5.1) as

$$\theta_{n+1}^\epsilon = \theta_n^\epsilon + \epsilon Y_n^\epsilon + \epsilon z_n^\epsilon,$$

where z_n^ϵ is the “correction term.” The decomposition (5.1) is the key to the analysis and first appeared in [25]. Under (3.2), the sequence $\{z_n^\epsilon, \epsilon > 0, n < \infty\}$ is uniformly integrable. Let q_ϵ be a sequence of integers satisfying (3.8'). The proof of Theorem 3.1 yields immediately that the limit of any convergent subsequence of $\{\theta^\epsilon(q_\epsilon \epsilon + \cdot), \epsilon > 0\}$ as $\epsilon \rightarrow 0$ has the form

$$(5.2) \quad \dot{\theta} = g(\theta) + z,$$

where $\int_0^t z(s)ds$ is the limit of the process with values $\epsilon \sum_{n=q_\epsilon}^{q_\epsilon+t/\epsilon-1} z_n^\epsilon$ for $t \geq 0$, and with the obvious change for $t < 0$. Thus the only problem concerns the characterization of $z(\cdot)$. By the uniform integrability, $(z(\cdot), \theta(\cdot))$ are Lipschitz continuous w.p.1 (Lemma 2.1). Clearly, $z(u) = 0$ on any interval $(t, t + \tau)$ in which $\theta(u) \in H^0$, the interior of H . To proceed, we need to specify H more fully, and we assume either of I or II below.

I. Let $q_i(\cdot), i = 1, \dots, p$, be continuously differentiable real-valued functions on R^r , with gradients $q_{i,x}(\cdot)$. Without loss of generality, let $q_{i,x}(x) \neq 0$ if $q_i(x) = 0$. Define $H = \{x : q_i(x) \leq 0, i = 1, \dots, p\}$ and assume that it is nonempty. Define $A(x)$, the set of active constraints at x , by $A(x) = \{i : q_i(x) = 0\}$. Define $C(x)$ to be the closed convex cone generated by $\{y : y = q_{i,x}(x), i \in A(x)\}$. Suppose that for each x with nonempty $A(x)$, the set $\{q_{i,x}(x), i \in A(x)\}$ is linearly independent.

II. H is an R^{r-1} -dimensional connected surface with a continuously differentiable outer normal. In this case, define $C(x), x \in H$, to be just the linear span of the outer normal at x .

THEOREM 5.1. *Assume the conditions above and the conditions of Theorem 3.1. Then the conclusions of Theorem 3.1 and Theorem 3.2 hold with (5.2) replacing (3.12) and $z \in -C(\theta(t))$, where the limit points of (5.2) replace M . If (3.11) is dropped, then the conclusions of Theorem 3.3 still hold. The same conclusions hold for the constrained form of Theorem 4.1.*

Proof. The basic proof is a straightforward extension of that of Theorem 3.1. To characterize $z(t)$ we use the fact that if for any (t, x) , $\theta^\epsilon(t) = x$, then $z_{t/\epsilon}^\epsilon$ is in a small neighborhood of $-C(y)$ for some y near x when ϵ is small. Then use the fact that $C(x)$ is upper semicontinuous in the sense that if $N_\delta(x)$ is a δ -neighborhood of x , then

$$\bigcap_{\delta>0} \bigcup_{y \in N_\delta(x)} C(y) \subset C(x);$$

i.e., the set of active constraints at x contains that for points very close to it. $\square$

Note. If, under I, there is only one active constraint (say, i) at t , and $g(\theta(t))$ points out of H , then the right-hand side of (5.2) is just the projection of $g(\theta(t))$ onto the boundary surface.

6. Applications of Theorem 3.1: Introduction. As they are stated, the results in §§3–5 do not explicitly deal with the optimization of an average cost over an infinite interval. In the examples in the remaining sections, we show that they are very powerful tools for proving convergence for just such problems. Three canonical examples of optimization will be described in detail. All use some approximation to a gradient search procedure. We will use constant step sizes as in Theorem 3.1, but the extensions to the decreasing step-size case will follow immediately from Theorem 4.1. Note that the constant step-size case $\epsilon_n = \epsilon$ has applications in tracking and adaptive control also. The examples concern the minimization of a stationary average cost associated with the path of a dynamical system. This section deals with a general discussion of the issues. Section 7 concerns a discrete event dynamical system example and an IPA-type estimator [13, 18]. Section 8 concerns a “piecewise deterministic” example, also using an IPA-type estimator and involving a problem in manufacturing. The third example involves a stochastic differential equations model. The examples are illustrative of many others using various methods of estimating derivatives.

Let us consider a canonical continuous time model in a rather informal way, since we wish only to illustrate the basic ideas in an unincumbered way. The general considerations hold also for discrete-time models, as will be seen in the next two sections. Among the points to be clarified is the so-called resetting of the IPA “accumulator.” It will be seen that it is often neither necessary nor desirable. The basic ideas are in [22, 27], but their full potential has not been realized in the literature.

Suppose that for fixed parameter θ , $x(\cdot, \theta)$ represents the dynamical state process of the system. In order to fix ideas, let $x(\cdot, \theta)$ be defined by the SDE

$$dx(t, \theta) = b(x(t, \theta), \theta)dt + dw.$$

For the sake of simple notation, let both x and θ be real valued and the function $b(\cdot)$ be smooth enough so that the following calculations make sense. We return to this example in a more thorough way in §9. For initial condition $x(0, \theta) = x(0)$, fixed parameter θ , and cost rate $c(\theta, x(s, \theta))$, define the average cost per unit time on $[0, T]$ by

$$C_T(\theta, x(0)) = E \frac{1}{T} \int_0^T c(\theta, x(s, \theta)) ds.$$

Suppose that $C_T(\theta, x(0))$ is continuously differentiable with respect to θ with gradient $C_{T,\theta}(\theta, x(0))$. Suppose that for each θ the limit $C(\theta) = \lim_T C_T(\theta, x(0))$ exists and does not depend on $x(0)$. Suppose that the pointwise limit of $C_{T,\theta}(\theta, x(0))$ exists and is denoted by $\hat{C}_\theta(\theta)$. Then $\hat{C}_\theta(\theta) = C_\theta(\theta)$.

We wish to use SA to minimize $C(\theta)$. A common procedure for updating θ via gradient search is based on the consistency of $C_{T,\theta}(\theta, x(0))$; i.e., it is a good estimator of $C_\theta(\theta)$ if T is large. Pursuing this idea, let us update the parameter at times $nT, n = 1, 2, \dots$, as follows. Letting θ_n^ϵ denote the n th choice of the parameter, use it on $[nT, nT + T)$ to get an estimator $\hat{Y}_n^\epsilon$ of $-C_{T,\theta}(\theta_n^\epsilon, x(\theta_n^\epsilon, nT))$. Then use

$$(6.1) \quad \theta_{n+1}^\epsilon = \theta_n^\epsilon + \epsilon \hat{Y}_n^\epsilon.$$

Let $x^\epsilon(\cdot)$ (with $x^\epsilon(0) = x(0)$) denote the actual physical state process with the time varying θ_n^ϵ used, i.e., on $[nT, nT + T)$ $x^\epsilon(t) = x(t, \theta_n^\epsilon)$ with the “initial condition” of $x(\cdot, \theta_n^\epsilon)$ at time nT being

$$(6.2) \quad x(nT, \theta_n^\epsilon) = x^\epsilon(nT).$$

Continuing, suppose that $\hat{Y}_n^\epsilon$ is an unbiased estimator of $-C_{T,\theta}(\theta_n^\epsilon, x^\epsilon(nT))$. This is equivalent to “restarting” the estimation procedure anew at each nT with initial condition $x^\epsilon(nT)$. To see what the limit ODE for (6.1) might be, proceed purely formally, let $\xi_n^\epsilon = x^\epsilon(nT)$ and apply Theorem 3.1 to get

$$(6.3) \quad \dot{\theta} = - \int C_{T,\theta}(\theta, \xi) \mu(d\xi|\theta).$$

The right side of (6.3) would not be close to $-C_\theta(\theta)$ unless (at least) T is large. For this reason, it is often suggested that T depend on either or both ϵ, n and go to infinity as one or both of these quantities goes to its limit. In [41], there are conditions for the convergence of the right side of (6.3) to $-C_\theta(\theta)$ for Markov chain models.

Before showing how to improve (6.3), let us look at a typical procedure more closely. In order to get a (pathwise) gradient estimator one generally introduces an auxiliary process $y(\cdot, \theta)$. For IPA estimators [13, 18, 32], this would be the pathwise derivative of $x(\cdot, \theta)$ with respect to θ ; for likelihood ratio estimators [32, 37, 38] this would be the score function which keeps the information on the derivative of the measure. Other methods such as smoothed perturbation analysis and rare perturbation analysis [4] use auxiliary information that represents the difference between the path $x(\cdot, \theta)$ and a perturbed one. See also the discussion of mean square derivatives and finite differences in §9.

For the model used in our illustrative example, the appropriate $y(\cdot, \theta)$ process is the mean square derivative defined by $x_\theta(t, \theta) = y(t, \theta)$:

$$\dot{y}(t, \theta) = b_x(x(t, \theta), \theta) + b_\theta(x(t, \theta), \theta)$$

with initial condition $y(0, \theta) = 0$. Define $z(\cdot, \theta) = (x(\cdot, \theta), y(\cdot, \theta))$. The estimator of $C_{T,\theta}(\theta, x(0))$ has the form

$$\frac{1}{T} \int_0^T \lambda(\theta, z(s, \theta)) ds,$$

where

$$\lambda(\theta, z(s, \theta)) = c_x(\theta, x(s, \theta))y(s, \theta) + c_\theta(\theta, x(s, \theta)).$$

Let $z^\epsilon(\cdot) = (x^\epsilon(\cdot), y^\epsilon(\cdot))$ be the actual process with the time-varying parameter used. Then the formal procedure leading to (6.3) would use

$$(6.4) \quad \hat{Y}_n^\epsilon = -\frac{1}{T} \int_{nT}^{nT+T} \lambda(\theta, z^\epsilon(s)) ds,$$

where on $[nT, nT + T)$ we have $y^\epsilon(t) = y(t, \theta_n^\epsilon)$ with initial condition $y^\epsilon(nT) = 0$.

Now consider an alternative where $x^\epsilon(\cdot)$ is as above but $y^\epsilon(\cdot)$ is not reset to zero at times nT . Use $y^\epsilon(0) = 0$ and on $[nT, nT + T)$ use $y^\epsilon(\cdot) = y(\cdot, \theta_n^\epsilon)$ with initial condition at nT defined recursively by

$$(6.5) \quad y^\epsilon(nT) = y(nT, \theta_n^\epsilon).$$

Then with the new definition of $z^\epsilon(\cdot)$, use the estimator

$$(6.6) \quad Y_n^\epsilon = -\frac{1}{T} \int_{nT}^{nT+T} \lambda(\theta_n^\epsilon, z^\epsilon(s)) ds.$$

Note that in general (6.6) would not be an unbiased estimator of $-C_{T,\theta}(\theta_n^\epsilon, x^\epsilon(nT))$ due to the “memory” in its “initial conditions.”

Now define the process $\xi_n^\epsilon = z^\epsilon(nT)$, and assume the conditions of Theorem 3.1. Then the ODE which characterizes the limit behavior is (3.12), where

$$G(\theta, \xi) = -E[Y_n^\epsilon | \xi_n^\epsilon = \xi, \theta_n^\epsilon = \theta].$$

By the definition of the invariant measure, we then have

$$(6.7) \quad g(\theta) \equiv \int G(\theta, \xi) \mu(d\xi | \theta) = \lim_n \frac{1}{n} \sum_{i=1}^n EG(\theta, \xi_i(\theta)),$$

where $\xi_i(\theta)$ is the stationary process under θ . Under either (3.11a) or (3.11b), the limit on the right side is the same if we used the process $\xi_n(\theta)$ with initial condition $\xi_0(\theta) = (x(0), 0)$. Thus, with this new initial condition (6.7) equals

$$(6.8) \quad -\lim \frac{1}{nT} E \int_0^{nT} \lambda(\theta, z(s, \theta)) ds = -\lim_T C_{T,\theta}(\theta, x(0)) = -C_\theta(\theta).$$

This is what we want since it yields the gradient descent ODE

$$(6.9) \quad \dot{\theta} = -C_\theta(\theta)$$

in lieu of the “biased” (6.3). In the parlance of the literature (e.g., [32]), (6.9) results when we do not reset the “accumulator.” While there has been some discussion of this preferable alternative, proofs and a clear understanding were lacking. In the next three sections, the details are filled in for three classes of applications.

7. A discrete example: A GI/G/1 queue. We consider the problem treated in [7, 11, 31, 32]. The model is a single-server queue with a renewal arrival process and general service time distribution, which is parametrized by $\theta > 0$. For notational simplicity, we suppose that θ is real valued, but the development and results are the same in general. For fixed θ let $X_i(\theta)$ denote the sojourn time of the i th customer and $K(\theta)$ be a bounded real-valued function with a continuous and bounded gradient. The cost of interest is

$$(7.1) \quad C(\theta) = \lim_N \frac{1}{N} \sum_{i=1}^N EX_i(\theta) + K(\theta) \equiv \hat{C}(\theta) + K(\theta),$$

and we wish to use SA to get the minimizing θ . Again, we suppose that the parameter θ is bounded. Indeed, the parameter might have to be restrained to some particular interval $[\theta_-, \theta_+]$ in order for the assumptions below to hold, and we assume that this is done. The example is widely studied, but the conditions used here are about as simple as one can expect. The structure of the problem is similar (from the point of view of SA) to those arising in other applications to single queues (and even for some network problems). For example, consider the multiclass problem [34], admission control [42], flow control in a closed network [43], routing in an open network [40], and routing in a closed network [17]. Appendix 4 discusses the decentralized case that is of interest in network models.

Fixed θ -process: Application of IPA. We proceed to make the usual assumptions to assure that $dX_i(\theta)/d\theta$ exists and can be estimated via IPA. Define the

parametrized service time distribution $F(\cdot|\theta)$, and suppose that it is weakly continuous in θ . Define the inverse function $F^{-1}(\cdot|\theta)$ by

$$F^{-1}(\chi|\theta) = \min\{\zeta : F(\zeta|\theta) \geq \chi\}, \quad \chi \in [0, 1],$$

and assume that it is differentiable in θ for each χ , with a bounded and continuous (uniformly in χ) derivative denoted by $F_{\theta}^{-1}(\chi|\theta)$. For fixed θ , let $\{\zeta_i(\theta), i < \infty\}$ denote the sequence of service times and define $\chi_n(\theta) = F(\zeta_n(\theta)|\theta)$ and the derivative $Z_n(\theta) = F_{\theta}^{-1}(\chi_n(\theta)|\theta)$. Let $Q_i(\theta)$ denote the queue length and $\tau_i(\theta)$ the (residual) time until the next arrival, all taken just after the departure of the i th customer. Then $x_n(\theta) = (Q_n(\theta), \tau_n(\theta))$ is a Markov process. Define the cost for the first N customers, initialized at an arbitrary initial condition, as

$$(7.2) \quad C_N(\theta, x_0) = \hat{C}_N(\theta, x_0) + K(\theta) = \frac{1}{N} \sum_{i=1}^N EX_i(\theta) + K(\theta).$$

Suppose that the busy periods have finite mean length for each fixed θ . Let $\bar{Z}_n(\theta)$ denote the sum of the $|Z_j(\theta)|$ in the n th busy period. Suppose that

$$(7.3) \quad \sup_{\theta} E|\bar{Z}_n(\theta)| < \infty.$$

Remark on (7.3). In many cases where IPA can be applied, the θ is a scale parameter of the service distribution. Then we have the form $F(\zeta|\theta) = F(\theta\zeta|1)$, $\zeta_n(\theta) = \theta F^{-1}(\chi_n(\theta)|1)$, and $Z_n(\theta) = \zeta_n(\theta)/\theta$, where the $\chi_n(\theta)$. Letting $N(\theta)$ denote the number of services in a busy period, Wald's identity yields $E|\bar{Z}_n(\theta)| = E\bar{Z}_n(\theta) = EN(\theta)EZ_j(\theta)$. If the system is stable, then $EN(\theta) < \infty$ and (7.3) holds.

Continuing, consider the estimator

$$(7.4) \quad \hat{Z}_m(\theta) = \frac{1}{m} \sum_{i=1}^m \sum_{j=v_i(\theta)}^i Z_j(\theta),$$

where $v_i(\theta)$ is the index of the first arrival in the busy period in which customer i arrives. If $Q_0 = 0$, then (7.4) is an unbiased estimator of the derivative of $\hat{C}_m(\theta, x_0)$, where x_0 is the state at time zero. It is an asymptotically consistent estimator in that [13] for each initial condition

$$(7.5) \quad E\hat{Z}_m(\theta) \rightarrow \hat{C}_{\theta}(\theta).$$

Henceforth, just to simplify notation and not have to worry about the possibly separate indices for arrivals and departures, we suppose that the queue starts empty. The conditions and results are the same in general.

The estimator for the SA. We wish to use SA to minimize the cost (7.1) via use of the IPA estimator. Suppose that the parameter is updated after the departure of each successive group of N customers. We use the "customer number" instead of real time. For fixed θ , the estimator used on the n th interval (the departures $[nN + 1, nN + N]$) is to be

$$\hat{Y}_n(\theta) = \frac{1}{N} \sum_{i=nN+1}^{nN+N} \sum_{j=v_i(\theta)}^i Z_j(\theta).$$

Recall that $v_{nN}(\theta)$ is the index of the first arrival in the busy period in which arrival (equivalently, departure) nN occurs. Let $D_n(\theta)$ denote the number of departures from the $(nN + 1)$ st to the end of the busy period in which departure nN occurs. It is zero if the nN th departure ends a busy period. Now, to separate the “past” from the contributions over $[nN + 1, nN + N)$, we split $\hat{Y}_n(\theta)$ by defining

$$\psi_n(\theta) = \sum_{j=v_{nN}(\theta)}^{nN} Z_j(\theta), \text{ “past”}$$

$$A_n(\theta) = \frac{1}{N} \sum_{i=nN+1}^{nN+N} \sum_{j=v_i(\theta) \vee (nN+1)-1}^i Z_j(\theta), \text{ “future,”}$$

$$B_n(\theta) = \frac{1}{N} \sum_{i=nN+1}^{nN+N} \sum_{j=v_i(\theta)}^{v_i(\theta) \vee (nN+1)-1} Z_j(\theta),$$

Then

$$\hat{Y}_n(\theta) = A_n(\theta) + B_n(\theta).$$

We can write

$$B_n(\theta) = D_n(\theta)\psi_n(\theta)/N.$$

Following the basic framework of §3, define the Markov chain $\xi_n(\theta) = (Q_n(\theta), \tau_{nN}(\theta), \psi_n(\theta))$. Define $G_i(\cdot)$ by

$$G_0(\theta, \xi_n(\theta)) = E_n A_n(\theta), \quad G_1(\theta, \xi_n(\theta)) = E_n D_n(\theta)/N,$$

where E_n denotes the expectation conditioned on all the systems data up to and including the time of the nN th departure. It is equivalent to conditioning on $\xi_n(\theta)$.

The functions $G_i(\cdot, \xi)$ are continuous in θ , uniformly in each compact (θ, ξ) set by the continuity assumptions made on the distribution of the service interval and the derivative of its inverse. In preparation for the conclusion of the SA argument, note that (and define $G(\cdot)$ by)

$$E_n \hat{Y}_n(\theta) = G_0(\theta, \xi_n(\theta)) + G_1(\theta, \xi_n(\theta))\psi_n(\theta) \equiv G(\theta, \xi_n(\theta))$$

and that

$$(7.6) \quad \lim_n \frac{1}{n} \sum_{i=1}^n EG(\theta, \xi_n(\theta)) = \lim_n \frac{1}{nN} E \sum_{i=1}^{nN} \sum_{j=v_i(\theta)}^i Z_j(\theta) = \hat{C}_\theta(\theta)$$

for each initial condition.

The SA problem. For the actual physical system with the time-varying parameter, let ζ_n^ϵ denote the actual service time of the n th customer and Z_n^ϵ the derivative of the inverse function, using the parameter θ_n^ϵ for $nN + 1 \leq j \leq nN + N$. Let v_i^ϵ be the index of the first arrival in the busy period in which customer i departs. Let

τ_i^ϵ be the residual time to the next arrival and Q_i^ϵ the queue length, all taken at the time of the i th departure. To update θ_n^ϵ we use the estimator

$$(7.7) \quad \hat{Y}_n^\epsilon = \frac{1}{N} \sum_{i=nN+1}^{nN+N} \sum_{j=v_i^\epsilon}^i Z_j^\epsilon.$$

The SA algorithm is

$$(7.8) \quad \theta_{n+1}^\epsilon = \left[\theta_n^\epsilon - \epsilon \hat{Y}_n^\epsilon - \epsilon K_\theta(\theta_n^\epsilon) \right]_{\theta_-}^{\theta_+}.$$

Now define

$$\psi_n^\epsilon = \sum_{j=v_{nN}^\epsilon}^{nN} Z_j^\epsilon,$$

and define $A_n^\epsilon, B_n^\epsilon$ analogously to what was done for the fixed θ case. Then $\hat{Y}_n^\epsilon = A_n^\epsilon + B_n^\epsilon \psi_n^\epsilon$ and $\xi_n^\epsilon = (Q_{nN}^\epsilon, \tau_{nN}^\epsilon, \psi_n^\epsilon)$ is a Markov chain. We can write

$$(7.9) \quad E_n^\epsilon \hat{Y}_n^\epsilon = G_0(\theta_n^\epsilon, \xi_n^\epsilon) + G_1(\theta_n^\epsilon, \xi_n^\epsilon) \psi_n^\epsilon = G(\theta_n^\epsilon, \xi_n^\epsilon).$$

In this example, the P_n^ϵ in (3.5) does not depend on ϵ, n . Thus, (3.6) holds. Assumption (3.7) follows from the assumptions on the service time distribution. Also, the G_n^ϵ in (3.9) does not depend on ϵ, n . Assumption (3.11b) follows from (7.6).

We need conditions which guarantee (3.2), (3.4), and (3.8). Define $\bar{Z}_n^\epsilon$ to be the sums of $|Z_j^\epsilon|$ over the n th busy period. Suppose that

$$(7.10) \quad \{\bar{Z}_n^\epsilon, \epsilon > 0, n < \infty\} \text{ is uniformly integrable,}$$

$$(7.11) \quad \sup_\theta E\zeta(\theta) < E[\text{interarrival time}].$$

By (7.10), (3.2) holds for $\hat{Y}_n^\epsilon$. The θ_n^ϵ are bounded and tightness of $\{\xi_n^\epsilon, \epsilon > 0, n < \infty\}$ follows from (7.3) and (7.11). Condition (3.8) follows from (7.3) and (7.11), and (3.10) is a consequence of (7.3). Now the convergence to the ODE

$$(7.12) \quad \dot{\theta} = -C_\theta(\theta) - K_\theta(\theta)$$

projected onto $[\theta_-, \theta_+]$ follows from Theorem 3.1. The case $\epsilon_n \rightarrow 0$ follows from Theorem 4.1.

Remark. We note that the updates need not be at regular intervals. If the interupdate times are bounded, then the end result is the same. Unbounded interupdate times might conceivably be of future interest, but at the moment it is hard to imagine an application in which it would be allowed, but with suitable conditions on the increments between updates we get the same result. If the estimator were reset at each nN , then the $\hat{Y}_n^\epsilon$ in (7.8) is replaced by A_n^ϵ and the $C_\theta(\theta)$ in (7.12) would be replaced by the biased quantity $\int \hat{C}_{N,\theta}(\theta, x) \mu_0(dx|\theta)$, where $\mu_0(\cdot|\theta)$ is the invariant measure of the $x_n(\theta)$ process defined above (7.2).

8. An example from manufacturing: A piecewise deterministic problem. We now consider an interesting example from [44]. The reference considers a manufacturing system with two unreliable tandem machines and is typical of many applications to production rate scheduling problems. Let $\alpha_i(t)$ denote the indicator function that machine i is working, and assume that these processes are independent renewal processes. The production rates $u_i(\cdot)$, $i = 1, 2$, of the machines can be controlled, subject to the machines' working and to upper bounds $\bar{u}_i$ on the rates. Machine 1 feeds into machine 2 via a buffer for surplus inventory, and the demand rate for the output of machine 2 is fixed at d . The dynamical state is the current inventory level $x(\cdot) = (x_1(\cdot), x_2(\cdot))$. The inventory of machine 2 can be negative (backlog). The reference assumes that the inventory process defined below satisfies a Harris recurrence condition, but we will not need to suppose that. The dynamical state equation is $\dot{x}(0) = 0$ and

$$\dot{x}_1(t) = u_1(t) - u_2(t), \quad \dot{x}_2(t) = u_2(t) - d.$$

The control problem is actually more conveniently formulated in terms of the surplus variables, defined by

$$s_1(t) = x_1(t) + x_2(t) = \int_0^t u_1(s) ds, \quad s_2(t) = x_2(t).$$

They consider control strategies of a threshold type on the $s_i(\cdot)$. (Their thresholds B_1, B_2 are the θ_1, θ_2 here.) In order to illustrate our method, we shall focus on one of their strategies, called the surplus control. Loosely speaking, for this control the production rate is held at the maximum value if the surplus is less than the threshold and tries to stay at the threshold if it ever reaches it. If (during the transient initial period) the surplus on some machine is higher than the threshold, then production on that machine is zero. For notational simplicity, we suppose that the initial surpluses do not exceed the thresholds. The surplus process $s_i(\cdot)$ evolves with deterministic slopes ($\bar{u}_i$, 0, or $-d$) which correspond to maximum production rate minus demand, production rate equals demand, and no production, and these change at random times which depend on the values of the renewal $\alpha(\cdot)$ and state processes $s(\cdot)$. With minor exceptions, the assumptions used here are implied by those in [44]. The $\alpha_i(\cdot)$ processes do not depend on the thresholds. It is assumed that the maximum production rates satisfy $\bar{u}_1 > \bar{u}_2 > d$. Also, it is supposed that $\theta_1 > \theta_2 \geq 0$, with each threshold subject to an upper bound, and that the SA algorithm is constructed to guarantee this.

DEFINITIONS (the fixed θ processes). Henceforth, for the fixed θ processes, we write the state and surplus variables as $x(\cdot, \theta)$, $s(\cdot, \theta)$, resp. For fixed thresholds θ , the control is such that the surplus processes are defined by

$$(8.1) \quad \dot{s}_1(t, \theta) = \begin{cases} (\bar{u}_1 - d)I_{\{s_1(t, \theta) < \theta_1\}} - dI_{\{s_1(t, \theta) > \theta_1\}} & \text{if } \alpha_1(t) = 1, \\ -d & \text{otherwise,} \end{cases}$$

$$(8.2) \quad \begin{aligned} \dot{s}_2(t, \theta) &= (\bar{u}_2 - d)I_{\{s_2(t, \theta) < \theta_2\}} - dI_{\{s_2(t, \theta) > \theta_2\}} \\ &\quad \text{if } \alpha_2(t) = 1 \text{ and } (s_1(t, \theta) \geq s_2(t, \theta) \text{ or } \alpha_1(t) = 1), \\ \dot{s}_2(t, \theta) &= -d \quad \text{otherwise.} \end{aligned}$$

The dynamics are such that $s_1(t, \theta) - s_2(t, \theta) \geq 0$ if this condition holds at time zero, which we suppose. (The condition will eventually hold in any case.)

The system just described is an example of a piecewise deterministic control system [8, 16], where $x(\cdot, \theta)$ is piecewise linear, with the intervals being random. In [44], the interval distributions are exponential so that the state process is Markovian. For $c_1 > 0, c_2^\pm > 0$, the cost rate of concern is

$$c(s) = c_1 s_1 + c_2^+ s_2^+ + c_2^- s_2^-.$$

Actually, the reference starts with $\bar{c}(x) = c_1 x_1 + c_2^+ x_2^+ + c_2^- x_2^-$, but in the derivative calculations switches to $c(\cdot)$. The two cost rates are equivalent, with appropriate definitions of the coefficients.

The thresholds are adjusted via an SA algorithm using IPA-type derivatives of the cost with respect to the θ_i , with the aim of minimizing the cost function

$$(8.3) \quad C(\theta) = \lim_M \frac{1}{M} \int_0^M c(s(t, \theta)) dt,$$

where we suppose that the limit exists for each θ in the desired range. The reference [44] derives auxiliary processes $y_j^{(i)}(t, \theta)$, which is the pathwise derivative of $s_j(t, \theta)$ with respect to θ_i . We now describe these derivative processes.

The (IPA) derivative processes. The following results are taken from the reference, with only the terminology changed to bring it into line with our own. Define the random time $\tau(\theta) = \inf\{t > 0 : s_1(t, \theta) = \theta_1\}$. Then

$$y_1^{(1)}(t, \theta) = I_{\{t \geq \tau(\theta)\}}.$$

Define the random times $\tau_i^k(\theta)$ recursively by $\tau_2^0(\theta) = \tau(\theta)$ and

$$\tau_1^k(\theta) = \min\{t \geq \tau_2^{k-1}(\theta) : s_1(t, \theta) = s_2(t, \theta)\},$$

$$\tau_2^k(\theta) = \min\{t \geq \tau_1^k(\theta) : s_2(t, \theta) = \theta_2\}.$$

Then the pathwise derivative of the surplus process at machine 2 with respect to θ_1 is

$$y_2^{(1)}(t, \theta) = \sum_{k=1}^{\infty} I_{\{\tau_1^k(\theta) \leq t \leq \tau_2^k(\theta)\}}.$$

Note that at most one of the indicator functions in the sum can be positive at a time. Clearly, the expression for $\dot{s}_1(t, \theta)$ implies that $y_1^{(2)}(t) = 0$. Define the additional random times $\gamma_i^k(\theta)$ recursively by $\gamma_2^0(\theta) = 0$ and

$$\gamma_1^k(\theta) = \min\{t \geq \gamma_2^{k-1}(\theta) : s_2(t, \theta) = \theta_2\},$$

$$\gamma_2^k(\theta) = \min\{t \geq \gamma_1^k(\theta) : s_1(t, \theta) = s_2(t, \theta)\}.$$

Then the pathwise derivative with respect to θ_2 of the surplus process at machine 2 is

$$y_2^{(2)}(t, \theta) = \sum_{k=1}^{\infty} I_{\{\gamma_1^k(\theta) \leq t \leq \gamma_2^k(\theta)\}}.$$

Again, at most one of the indicator functions can be positive at a time.

Let $z(\cdot, \theta) = (x(\cdot, \theta), y(\cdot, \theta))$. Define $Y_n(\theta) = (Y_n^1(\theta), Y_n^2(\theta))$, where

$$(8.4) \quad Y_n^i(\theta) = -\frac{1}{T} \int_{nT}^{nT+T} \lambda_i(\theta, z(t, \theta)) dt,$$

where

$$(8.5a) \quad \lambda_1(\theta, z(s, \theta)) = c_1 y_1^{(1)}(t, \theta) + c_2^+ y_2^{(1)}(t, \theta) I_{\{s_2(t, \theta) \geq 0\}} - c_2^- y_2^{(1)}(t, \theta) I_{\{s_2(t, \theta) < 0\}},$$

$$(8.5b) \quad \lambda_2(\theta, z(s, \theta)) = c_2^+ y_2^{(2)}(t, \theta) I_{\{s_2(t, \theta) \geq 0\}} - c_2^- y_2^{(2)}(t, \theta) I_{\{s_2(t, \theta) < 0\}}.$$

Then $Y_n^i(\theta)$ is an unbiased estimator for $-C_{T, \theta_i}(\theta, x(0))$.

The SA updates will be at times $nT, n = 1, \dots$, for some $T > 0$. Below, we will be concerned with the set $(s(nT, \theta), y(nT, \theta), \alpha(nT))$. For a general renewal process, this set is not a Markov process. To Markovianize, we augment it by adding the residual times until the next change of values of the $\alpha_1(\cdot), \alpha_2(\cdot)$ after nT . Let $\xi_n(\theta)$ denote the consequent quadruple.

To minimize work, we suppose that for each initial condition the limit

$$(8.6) \quad -\lim_M \frac{1}{M} \int_0^M E \lambda(\theta, z(\theta, t)) dt$$

exists for each θ value of interest and is continuous in θ . Define it as $g(\theta)$. These conditions are weaker than those in the reference. Then $C_\theta(\theta) = -g(\theta)$. In any case, these conditions amount to nothing more than asymptotic consistency, and are a minimal condition for the convergence. Define $G(\xi, \theta) = E[Y_n(\theta) | \xi_n(\theta) = \xi]$. Then $G(\cdot)$ is continuous and bounded. The limit (8.6) is the same as

$$\lim_n \frac{1}{n} \sum_{i=1}^n EG(\xi_n(\theta), \theta).$$

In order to prove the tightness of $\{s^\epsilon(t)\}$, we need the following conditions. Let p_i be the stationary probability that machine i is working. We suppose that

$$(8.7) \quad \bar{u}_2 p_2 > \bar{u}_1 p_1 > d;$$

i.e., the average maximum possible production rate for machine 2 is greater than that of machine 1, which is greater than the demand rate. Also, suppose that, where E_t is the expectation given $\{\alpha_i(v), v \leq t, i = 1, 2\}$,

$$(8.8) \quad \int_t^\infty E_t[\alpha_i(v) - p_i] dv = O(1),$$

where $O(1)$ means that the term is bounded uniformly in all variables. Loosely speaking, (8.8) is equivalent to the expectation of the time to the next change in the $\alpha_i(\cdot)$ being uniformly bounded, conditioned on the current data. This is certainly not a strong condition.

THE SA ALGORITHM. Fix $T > 0$, the time interval between parameter updates. Let $s^\epsilon(\cdot)$ denote the actual surplus process with the time-varying parameter. Define $y^\epsilon(\cdot)$ as the derivative process with the random times determined by the actual time of

the associated events in the true physical process. It is not reset at each nT . Then set $z^\epsilon(\cdot) = (x^\epsilon(\cdot), z^\epsilon(\cdot))$. Let Y_n^ϵ be (8.4) with $z^\epsilon(\cdot)$ used in lieu of $z(\cdot, \theta)$. The algorithm is now (3.1).

Convergence of the SA. Condition (3.2) holds since the Y_n^ϵ are uniformly bounded. Tightness of all the components of the Markov chain (as needed for (3.4)) follows once tightness of $\{s^\epsilon(nT), \epsilon > 0, n < \infty\}$ is shown. This will be discussed at the end of the section. The $P(\cdot)$ and $G(\cdot)$ do not depend on n or ϵ . The weak continuity of the transition probability is a consequence of the basic structure of the problem. In particular, of the continuous effects of the threshold variations and the monotone nature of the evolution of the residual times. Finally, (3.11b) holds by assumption (8.6). Thus, the conclusions of Theorem 3.1 hold, and the extensions of Theorem 3.1 can also be readily handled. Theorem 3.1 asserts that the limit ODE is $\dot{\theta} = g(\theta)$ for the function $g(\theta)$ defined above. The reference [44] presents numerical data which implies that the cost function has a unique minimum and that their SA converges nicely.

The requirements are generally much weaker than those in the reference, and we do not need to restart the estimator periodically or let $T \rightarrow \infty$ as $\epsilon \rightarrow 0$. Some other references concerned with the use of SA in related manufacturing problems are [6, 39, 16, 45]. In [16], another interesting work on the same subject, they use an SA with gains $\epsilon_n \rightarrow 0$ and an IPA-type estimator where the estimation intervals go to infinity as $n \rightarrow \infty$. They do not “reset the accumulator.” The conditions used here are simpler whether or not the step size is constant. The paper [19] was one of the early works which attempted to improve the operation of a production line subject to random breakdowns using IPA-type estimates, and dealt with a production line in an automobile factory. Some of the background analytical work is in [20].

Tightness of $\{s^\epsilon(t)$, small $\epsilon > 0, t < \infty\}$. Let $\mathcal{B}_t$ denote the minimal sigma-algebra measuring $\{\alpha_i(v), v \leq t, i = 1, 2\}$ and E_t the associated conditional expectation. We define a differential operator $\tilde{A}$ and its domain.

The real-valued functions $f(\cdot), g(\cdot)$ of (t, ω) will be measurable with $f(t), g(t)$ being $\mathcal{B}_t$ -measurable. Suppose that for each $T < \infty$,

$$\sup_{t \leq T} E|g(t)| < \infty, \limsup_{\delta \downarrow 0} \sup_{t \leq T} E \left| \frac{E_t f(t + \delta) - f(t)}{\delta} \right| < \infty,$$

$$\lim_{\delta \downarrow 0} E \left| \frac{E_t f(t + \delta) - f(t)}{\delta} - g(t) \right| = 0.$$

Then we say that $\tilde{A}f(\cdot) = g(\cdot)$. The process $f(t) - \int_0^t g(v)dv$ is a martingale [21], [22, §3.2.2].

The $s_i^\epsilon(t)$ are bounded above by the upper bounds to the thresholds. Thus the tightness problem concerns the probability of large negative excursions. We will work with altered processes, which provide the appropriate bounds from below. First we work with $s_1^\epsilon(t)$. To get a lower bound, we can suppose that $\theta_{i,n}^\epsilon = 0$. Let $q_1(t)$ be the process (8.1) with $\theta_1 = 0$. Then $\tilde{A}q_1^2(t)^2/2 = q_1(t)\dot{q}_1(t)$, which is $q_1(t)[\bar{u}_1\alpha_1(t) - d]$. To help in averaging the term with the $\alpha_1(t)$, define

$$\tilde{q}_1(t) = q_1(t)\bar{u}_1 \int_t^\infty E_t[\alpha_1(v) - p_1]dv.$$

We have

$$\tilde{A}\tilde{q}_1(t) = -q_1(t)\bar{u}_1[\alpha_1(t) - p_1] + O(1).$$

Define $Q_1(t) = q_1^2(t)/2 + \tilde{q}_1(t)$. Then

$$(8.9) \quad \tilde{A}Q_1(t) = q_1(t)[\bar{u}_1 p_1 - d] + O(1).$$

By (8.7), $\bar{u}_1 p_1 - d > 0$. Thus there are $k_i > 0$ such that for $q_1 \leq -k_1$, we have the right side of (8.9) less than $-k_2$. This implies that when $q_1 \leq -k_1$, $Q_1(\cdot)$ has the supermartingale property (until it hits the interval $[-k_1, 0]$). These considerations and the quadratic dependence of $Q_1(t)$ on $q_1(t)$ imply the tightness of $\{Q_1(t), t < \infty\}$. The tightness of $\{q_1(t), t < \infty\}$ (hence of $\{s_1^\epsilon(t), \text{small } \epsilon > 0, t < \infty\}$) follows from this tightness and the quadratic dependence of $Q_1(t)$ on $q_1(t)$.

The tightness of the $\{s_2^\epsilon(t)\}$ is proved in the same way. By the above results, it is sufficient to prove tightness for $s_1^\epsilon(t) - s_2^\epsilon(t)$ instead. Again, this can be done by a bounding argument. We have $s_1^\epsilon(t) - s_2^\epsilon(t) \geq 0$. Thus, we need to be concerned with large positive excursions of this difference. We start by fixing the thresholds at the upper bound for s_1^ϵ and the lower bound for s_2^ϵ . Once these thresholds are fixed, their actual values do not affect the result, so we can set them equal to zero without loss of generality. Let $q_i(\cdot)$ denote the new processes with the thresholds fixed at zero. One starts the argument by using a tentative Liapunov function (for the variables with the thresholds fixed at zero): $(q_1(t) - q_2(t))^2/2$. One bounds the derivative from above. Then introduces a function $\tilde{q}_2(\cdot)$ whose purpose is analogous to that of $\tilde{q}_1(\cdot)$ above. We omit the rest of the details due to lack of space. But by an argument similar to what was done for $q_1(\cdot)$ above, we get the tightness under the conditions (8.7), (8.8).

9. A continuous time SDE example: The system. We continue the discussion of the SDE model of §6 but with more detail and a more general system. We start by using the mean square derivatives and then discuss finite-difference forms. The finite-difference forms can be advantageous. One can use them without knowing the exact model and for more general cost functions. They can also be used for discrete-event systems in the same way. Let θ be real valued (for notational simplicity only) and $x \in R^k$. Let $b(\cdot)$ be a R^k -valued and continuously differentiable function of (x, θ) with bounded x and θ first derivatives, $\sigma(\cdot)$ a continuously differentiable matrix-valued function of x with bounded first derivatives, and let the fixed θ state process satisfy the SDE

$$(9.1) \quad dx(t, \theta) = b(x(t, \theta), \theta)dt + \sigma(x(t, \theta))dw(t),$$

where $w(t)$ is a standard vector valued Wiener process. Define the auxiliary process $y(t, \theta)$ by

$$(9.2) \quad dy(t, \theta) = b_x(x(t, \theta), \theta)y(t, \theta)dt + b_\theta(x(t, \theta), \theta)dt + (\sigma, y)(t, \theta)dw(t),$$

where the vector $(\sigma, y)(t, \theta)dw(t)$ is defined by its components

$$\sum_{j,p} \frac{\partial \sigma_{ij}(x(t, \theta))}{\partial x_p} y_p(t, \theta) dw_j(t), \quad i = 1, \dots, k,$$

The $y(t, \theta)$ is the pathwise (mean square) derivative of $x(t, \theta)$ with respect to θ . This “pathwise derivative” for the SDE was in use [12] long before its analogue for the

discrete case was developed. Define $z(\cdot, \theta) = (x(\cdot, \theta), y(\cdot, \theta))$. Let $c(\cdot, \cdot)$ be a bounded, real-valued, continuously differentiable function of (θ, x) with bounded derivatives, and define $C_T(\theta) = \int_0^T c(\theta, x(s, \theta)) ds/T$ as in §6.

The SA procedure. Use the method of §6, where we update at intervals $nT, n = 1, \dots$, with θ_n^ϵ being the parameter value used on $[nT, nT + T)$. Use (3.1) with $x^\epsilon(\cdot)$ again defined as the state process with the time-varying parameter used. Define $y^\epsilon(\cdot)$ as above (6.5) (i.e., it is never reset), and define $z^\epsilon(\cdot) = (x^\epsilon(\cdot), y^\epsilon(\cdot))$. Define

$$(9.3) \quad Y_n^\epsilon = -\frac{1}{T} \int_{nT}^{nT+T} \left[\sum_j c_{x_j}(\theta_n^\epsilon, x^\epsilon(s)) y_j^\epsilon(s) + c_\theta(\theta_n^\epsilon, x^\epsilon(s)) \right] ds.$$

We assume the following conditions.

$$(9.4) \quad \text{The process } \{z(nT, \theta)\} \text{ has a unique invariant measure for each } \theta.$$

$$(9.5) \quad \{z^\epsilon(nT), \theta_n^\epsilon, \epsilon > 0, n < \infty\} \text{ is tight}$$

$$(9.6) \quad \{Y_n^\epsilon, \epsilon > 0, n < \infty\} \text{ is uniformly integrable.}$$

$$(9.7) \quad \{z(0, \theta) : \theta \in \Theta \text{ compact}, z(\cdot, \theta) \text{ stationary}\} \text{ is tight.}$$

Condition (9.4) implies that the limit $C(\theta)$ of $C_T(\theta, x(0))$ exists and does not depend on $x(0)$. Under these conditions, Theorem 3.1 and its extensions hold. Thus (6.9) holds for algorithm (3.1). An SA procedure using mean square derivatives was used to good practical effect in [5, 26]. There is an analogous result under (3.11b).

Finite-difference methods. The main difficulties in applications concern the verification of the various conditions on the y processes. This was an unresolved issue in [5]. These difficulties can be alleviated by using a finite-difference method rather than the derivative $y^\epsilon(\cdot)$ process. We will discuss two forms of the finite-difference method. The first is the more traditional, using separate runs for the different components of the difference. The second combines these runs into one “concatenated difference” and provides a useful alternative since it can be used on line. There is an obvious analogue for discrete-event systems.

A finite-difference alternative: Simultaneous runs. Tightness and uniqueness of the appropriate invariant measure are often much easier to prove if a finite-difference method is used in lieu of the estimator (9.3), since then the troublesome $y^\epsilon(\cdot)$ process does not appear. We retain the conditions of the last part, with the exception of those concerning the y process. We also let $c(\cdot)$ be simply bounded and continuous. Given a finite-difference interval $\delta\theta$, replace the integrand in (9.3) with

$$(9.8) \quad \frac{c(\theta_n^\epsilon + \delta\theta, x(s, \theta_n^\epsilon + \delta\theta)) - c(\theta_n^\epsilon - \delta\theta, x(s, \theta_n^\epsilon - \delta\theta))}{2(\delta\theta)}.$$

Here we use two separate simulations, one for $\{\theta_n^\epsilon + \delta\theta\}$ and one for $\{\theta_n^\epsilon - \delta\theta\}$. We thus run two processes $x^{\epsilon, \pm}(\cdot)$ defined by $x^{\epsilon, \pm}(0) = x(0)$, and on $[nT, nT + T)$ set $x^{\epsilon, \pm}(\cdot) = x(\cdot, \theta_n^\epsilon \pm \delta\theta)$ with initial condition at nT defined recursively by $x(nT, \theta_n^\epsilon \pm \delta\theta) = x^{\epsilon, \pm}(nT)$. Generally, one would want to use the same Wiener process to drive

the two processes. This (common random variables) form often yields essentially the same path properties as does the use of the derivative process.

Under the given conditions, Theorem 3.1 yields that the limit ODE is

$$(9.9) \quad \dot{\theta} = -\frac{1}{2\delta\theta} \int [c(\theta + \delta\theta, \xi)\mu(d\xi|\theta + \delta\theta) - c(\theta - \delta\theta, \xi)\mu(d\xi|\theta - \delta\theta)],$$

where $\mu(\cdot|\theta)$ is the invariant measure of $\{x(nT, \theta)\}$, and with the analogous formula for the multidimensional θ case. Due to the additive way that the two terms appear in (9.8), we do not need to have a unique invariant measure of the pair $\{x(nT, \theta + \delta\theta), x(nT, \theta - \delta\theta)\}$ for each θ but only of $\{x(nT, \theta)\}$ for each θ .

The finite-difference approach can be either easier or harder than the pathwise derivative approach. The order of the SDEs to be solved in each case is the same. If $\sigma(x)$ actually depended on x , then the pathwise derivative procedure cannot be conducted “on line,” since we need to know the Wiener process to get $y(\cdot, \theta)$. If $\sigma(x)$ does not depend on x , then the equation for $y(\cdot, \theta)$ or $y^\epsilon(\cdot)$ is linear in the y variable (but with time varying coefficients) and it is simpler to solve. The procedure can then be done “on line,” at least in principle. An additional point to be kept in mind is that any simulation can only approximate the solution to (9.1) and (9.2). Thus, there is the additional question concerning the relations between the estimators for the approximations and those of the original model. See [26] for some results on this important problem. Finally, the finite-difference method can be used for cases where the $c(\cdot), b(\cdot)$ are not smooth, e.g., where $c(\cdot)$ is an indicator function of a event of interest.

Finite differences with only one run. Alternatively to the traditional simultaneous run method discussed above, a single run can be used to get a good estimate of the desired quantity and will be useful when the optimization must be done “on line,” where simultaneous runs might not be possible. Let $T > 0$ and $\delta\theta > 0$ be given. For the “one run” method, we use $\theta_n^\epsilon + \delta\theta$ on the interval $[2nT, 2nT + T)$ and then $\theta_n^\epsilon - \delta\theta$ on $[2nT + T, 2nT + 2T)$. Let $x^\epsilon(\cdot)$ denote the actual process with the $\theta_n^\epsilon \pm \delta\theta$ being used on the appropriate alternating time intervals. The appropriate fixed θ process, which we call $\hat{x}(\cdot, \theta)$, uses parameter value $\theta + \delta\theta$ on $[0, T)$ and then alternates between $\theta - \delta\theta$ and $\theta + \delta\theta$ on successive intervals of width T . We use

$$Y_n^\epsilon = -\frac{1}{2T\delta\theta} \int_0^T [c(\theta_n^\epsilon + \delta\theta, x^\epsilon(nT + s)) - c(\theta_n^\epsilon - \delta\theta, x^\epsilon(nT + T + s))] ds.$$

The analysis follows the lines of Theorem 3.1, but the limit form will be slightly different from that above. It is worth commenting on the differences between the simultaneous and single run cases since they are of practical importance and of interest in related algorithms. The main additional problem is due to the fact that the transition function for the fixed θ process depends periodically on time.

Let $\xi_n^+(\theta) = \hat{x}(2nT, \theta)$ and $\xi_n^-(\theta) = \hat{x}(2nT + T, \theta)$. Suppose that the stationary processes exist and are unique, with invariant measures $\mu^+(\cdot|\theta)$ and $\mu^-(\cdot|\theta)$, resp. Define

$$G^+(\theta, \xi) = \frac{1}{2T\delta\theta} \int_0^T E[c(\theta + \delta\theta, x(s, \theta + \delta\theta)) | x(0) = \xi] ds,$$

$$G^-(\theta, \xi) = \frac{1}{2T\delta\theta} \int_0^T E[c(\theta - \delta\theta, x(s, \theta - \delta\theta)) | x(0) = \xi] ds.$$

The right side of the limit ODE is

$$(9.10) \quad g(\theta) = - \int [G^+(\theta, \xi) \mu^+(d\xi|\theta) - G^-(\theta, \xi) \mu^-(d\xi|\theta)].$$

Let $P_T(\xi, \cdot|\theta + \delta\theta)$ denote the transition function for the process $x(nT, \theta + \delta\theta)$. Note that

$$(9.11) \quad \mu^-(d\xi|\theta) = \int \mu^+(d\tilde{\xi}|\theta) P_T(\tilde{\xi}, d\xi|\theta + \delta\theta).$$

Thus, as $\delta\theta \rightarrow 0$, the $\mu^\pm(\cdot|\theta)$ converge weakly to $\mu(\cdot|\theta)$, and so do the $\mu(\cdot|\theta \pm \delta\theta)$. Thus the $\mu^\pm(\cdot|\theta)$ become closer to the $\mu(\cdot|\theta \pm \delta\theta)$, which are the measures in the right side of (9.9). This line of reasoning suggests that the one sample procedure might be quite reasonable. The obvious form of (3.11b) can replace the assumption of uniqueness of the invariant measures.

To better understand the above “one-run” procedure, one needs to compare it to an alternative one-run procedure, say where we restart the process each T units of time at some fixed initial value, still using the $\theta \pm \delta\theta$ on the alternate intervals (assuming that such restarts were possible in the application). This would yield a right side of the form (9.10), where the $\mu^\pm$ are replaced by the measures concentrated on the fixed initial values. We expect that this “restarted method” would be much inferior to the original procedure, since the $\mu^\pm(\cdot|\theta)$ defined above would be much closer to the desired values $\mu(\cdot|\theta \pm \delta\theta)$, particularly for large T . The situation would be a little more complicated if θ were vector valued, but the general idea is the same. Analogous remarks can be made on the use of finite differences for discrete-event systems.

Appendix 1. Non-Markov models. Consider the algorithm (3.1). Suppose that due to the nature of the correlations, there is no convenient Markov chain $\{\xi_n^\epsilon, n < \infty\}$ for each ϵ . For example, the service or interarrival intervals in a queue might be correlated in a “non-Markovian way.” The first-order perturbed test function methods of [22] are often very helpful in such circumstances, and we will outline the general idea in the context of Theorem 3.1.

For each $\epsilon > 0$, $\{Y_n^\epsilon, n < \infty\}$ denotes the observation sequence, and the uniform integrability (3.2) is assumed. The θ_n will be assumed to be in a compact set to make the development simpler. For fixed parameter θ and each integer m we define the fixed θ process $\{Y_j^m(\theta), j \geq m\}$, and define $Y_j^m(\theta) = Y_j^\epsilon$ for $j \leq m$ by supposing that after time m the sequence evolves as though the parameter were held fixed at θ . This process is the analogue of the fixed θ Markov chain of §3. The key to the development is to work with an appropriately chosen “perturbed” θ_n^ϵ , which differs only slightly from θ_n^ϵ and for which the theorem can be proved. Suppose that there is a continuous function $g(\cdot)$ such that for each large $T_1 < T_2 < \infty$ and $m < T_1/\epsilon$, the sum defined by

$$(A1.1) \quad \delta f_m^\epsilon(\theta) = \sum_{j=m}^{T_2/\epsilon} \epsilon E_m^\epsilon [Y_j^m(\theta) - g(\theta)]$$

goes to zero in mean, uniformly in $m \leq T_1/\epsilon$ as $\epsilon \rightarrow 0$. The convergence of (A1.1) is a condition on the “mixing rate” of the noise process.

Define $\delta f_n^\epsilon = \delta f_n^\epsilon(\theta_n^\epsilon)$. In the analysis, $\tilde{\theta}_n^\epsilon = \theta_n^\epsilon + \delta f_n^\epsilon(\theta_n^\epsilon)$ replaces θ_n^ϵ . We also need the continuity condition that

$$(A1.2) \quad \sum_{j=m}^{T_2/\epsilon} E_m^\epsilon [Y_j^m(\theta + \delta\theta) - g(\theta + \delta\theta)] - \sum_{j=m}^{T_2/\epsilon} E_m^\epsilon [Y_j^m(\theta) - g(\theta)] \rightarrow 0$$

in the mean, uniformly in $m \leq T_1/\epsilon$ as $\delta\theta \rightarrow 0$.

We have the following theorem.

THEOREM A1.1. *Let q_ϵ satisfy (3.8'). Then, assuming (3.2) and the conditions concerning (A1.1) and (A1.2), $\{\theta^\epsilon(\epsilon q_\epsilon + \cdot), \epsilon > 0\}$ is tight, and the limit of any weakly convergent subsequence satisfies (3.12). If $\epsilon q_\epsilon \rightarrow \infty$, then the conclusions of Theorem 3.2 hold.*

Proof. The proof is much simpler than that of Theorem 3.1. Again, for simplicity, we let $q_\epsilon = 0$. Let $\tilde{\theta}^\epsilon(\cdot)$ denote the continuous parameter interpolation (interval ϵ) of the $\tilde{\theta}_n^\epsilon$ sequence defined above (A1.2). $\{\theta^\epsilon(\cdot)\}$ is tight. For notational simplicity, let ϵ index a weakly convergent subsequence. Let $h(\cdot), s_i, t, \tau$ be as in Theorem 3.1 with $t + \tau \leq T_1$, and suppose for notational simplicity that ϵ indexes a weakly convergent subsequence. By the definition of conditional expectation,

$$(A1.3) \quad Eh(\tilde{\theta}^\epsilon(s_i), i \leq q) \left[\tilde{\theta}^\epsilon(t + \tau) - \tilde{\theta}^\epsilon(t) - \sum_{m=t/\epsilon}^{(t+\tau)/\epsilon-1} E_m^\epsilon (\tilde{\theta}_{m+1}^\epsilon - \tilde{\theta}_m^\epsilon) \right] = 0.$$

We have

$$(A1.4) \quad E_m^\epsilon (\tilde{\theta}_{m+1}^\epsilon - \tilde{\theta}_m^\epsilon) = \epsilon E_m^\epsilon Y_m^\epsilon + E_m^\epsilon [\delta f_{m+1}^\epsilon - \delta f_m^\epsilon].$$

The last term on the right equals

$$\epsilon [g(\theta_m^\epsilon) - E_m^\epsilon Y_m^m(\theta_m^\epsilon)] + \epsilon W_m^\epsilon,$$

where

$$W_m^\epsilon = \sum_{j=m+1}^{T_2/\epsilon} E_m^\epsilon [Y_j^{m+1}(\theta_{m+1}^\epsilon) - g(\theta_{m+1}^\epsilon)] - \sum_{j=m+1}^{T_2/\epsilon} E_m^\epsilon [Y_j^m(\theta_m^\epsilon) - g(\theta_m^\epsilon)].$$

Hence, we can write (A1.4) as $\epsilon g(\theta_m^\epsilon) + \epsilon W_m^\epsilon$. By definition,

$$Y_j^{m+1}(\theta_{m+1}^\epsilon) = Y_j^m(\theta_{m+1}^\epsilon), \quad j \geq m+1.$$

Therefore, if the Y_n^ϵ were bounded (so that $|\theta_{m+1}^\epsilon - \theta_m^\epsilon| \rightarrow 0$ as $\epsilon \rightarrow 0$ uniformly in (m, ω)), we could use (A1.2) to get that $E|W_m^\epsilon| \rightarrow 0$ uniformly in $m : m\epsilon \leq T_1$ as $\epsilon \rightarrow 0$. Then, (A1.3) would imply that

$$(A1.5) \quad \lim_{\epsilon} Eh(\theta^\epsilon(s_i), i \leq q) \left[\theta^\epsilon(t + \tau) - \theta^\epsilon(t) - \epsilon \sum_{m=t/\epsilon}^{(t+\tau)/\epsilon-1} g(\theta_m^\epsilon) \right] = 0.$$

The theorem would follow from this last equality, analogously to the situation in Theorem 3.1. If the Y_n^ϵ are not bounded, use the uniform integrability (3.2) to bound them for the purposes of the proof. For $B > 0$ define $Y_{n,,B}^\epsilon$ to be Y_n^ϵ but with

the components truncated at $\pm B$. Define $\hat{\theta}_m^\epsilon$ as follows. Let $\hat{\theta}_{t/\epsilon}^\epsilon = \theta_{t/\epsilon}^\epsilon$. Then, for $(t + \tau)/\epsilon \geq m \geq t/\epsilon$, set $\hat{\theta}_{m+1}^\epsilon = \hat{\theta}_m^\epsilon + \epsilon Y_{m,B}^\epsilon$. Now proceed with $\hat{\theta}_n^\epsilon$ replacing θ_n^ϵ , but continuing to use the original definition of W_n^ϵ . The result is (A1.5) plus an error which goes to zero as $B \rightarrow \infty$. $\square$

An interpretation. Refer to the example in §7. Fix the parameter at θ . Suppose that $Y_n(\theta)$ is the IPA estimator on the interval $[nN, nN + N)$ without “resetting the accumulator.” Then

$$Y_0(\theta) + \cdots + Y_{n-1}(\theta)$$

is an unbiased estimator of the derivative of the cost on $[0, nN]$. Suppose that for each fixed θ , the system is stationary. The condition (A1.1) is close to the assumption that

$$(A1.6) \quad \frac{1}{n} \sum_{i=0}^{n-1} EY_i(\theta) \rightarrow g(\theta)$$

for each initial condition, the only difference being in the conditioning data. Suppose that the mean cost per unit time on $[0, T]$ converges as time goes to infinity. This convergence and the convergence of the mean value of the left side of (A1.6) to $g(\theta)$ imply (the closed graph theorem) that $g(\theta)$ is the derivative of the mean ergodic cost at θ . Analogous comments apply to the example of §8.

The extensions of Theorem A1.1 are handled analogously to the way that the extensions of Theorem 3.1 were handled; e.g., for the analog of Theorem 3.3, replace the T_2/ϵ in (A1.1) by $m_n(T_2)$ and ϵ by ϵ_n . The general scheme is very flexible and allows many variations. More background and examples satisfying the conditions is in [22].

Appendix 2. An alternative averaging method. Return to (3.17) and the problem of replacing $(\xi_j^\epsilon, \theta_j^\epsilon)$ by $(\xi_j(\theta_{l_n}^\epsilon), \theta_{l_n}^\epsilon)$ in

$$(A2.1) \quad \frac{1}{n_\epsilon} \sum_{j=l_{n_\epsilon}}^{l_{n_\epsilon}+n_\epsilon-1} E_{l_{n_\epsilon}}^\epsilon G_j^\epsilon(\theta_j^\epsilon, \xi_j^\epsilon) = \frac{1}{n_\epsilon} \sum_{j=0}^{n_\epsilon-1} E_{l_{n_\epsilon}}^\epsilon G_{l_{n_\epsilon}+j}^\epsilon(\theta_{l_{n_\epsilon}+j}^\epsilon, \xi_{l_{n_\epsilon}+j}^\epsilon).$$

We present an alternative approach which avoids the use of the occupation measure $R(l, \epsilon, \cdot)$ but involves some other conditions. The method relies more heavily on continuity properties. Only a brief outline will be given, but the main idea should be clear. The sequence of integers n_ϵ might be different here than in Theorem 3.1. Continue to assume (3.2)–(3.5), (3.8), and (3.10). Suppose that $P_n^\epsilon(\xi, \cdot | \theta)$ is weakly continuous in (θ, ξ) , uniformly in (ϵ, n) and in each compact (θ, ξ) set. Let $G_n^\epsilon(\theta, \xi)$ be continuous in (θ, ξ) , uniformly in (ϵ, n) and in each compact (θ, ξ) set.

Due to the tightness (3.4), the uniform integrability (3.2) and the assumed uniform θ continuity of the $G_j^\epsilon(\cdot)$, we can suppose (as in Theorem 3.1) that the Y_n^ϵ are truncated and that in the interval $[t, t + \tau]$ of concern (see proof of Theorem 3.1) the θ_n^ϵ take values in some compact set. Thus we can suppose that $|\theta_{j+1}^\epsilon - \theta_j^\epsilon| = O(\epsilon)$ for all j of interest. For notational simplicity, we will not use the truncation notation.

Define

$$P_n^\epsilon\{d\xi, d\theta | \xi_n^\epsilon, \theta_n^\epsilon\} = P_n^\epsilon\{\xi_{n+1}^\epsilon \in d\xi, \theta_{n+1}^\epsilon \in d\theta | \xi_n^\epsilon, \theta_n^\epsilon\}.$$

Writing out the conditional expectation $E_{l_{n_\epsilon}}^\epsilon G_{l_{n_\epsilon}+j}^\epsilon(\theta_{l_{n_\epsilon}+j}^\epsilon, \xi_{l_{n_\epsilon}+j}^\epsilon)$ in (A2.1), we have

$$\int \cdots \int P_{l_{n_\epsilon}}^\epsilon \{d\xi_1, d\theta_1 | \xi_{l_{n_\epsilon}}^\epsilon, \theta_{l_{n_\epsilon}}^\epsilon\} \cdots P_{l_{n_\epsilon}+j-1}^\epsilon \{d\xi_j, d\theta_j | \xi_{j-1}, \theta_{j-1}\} G_{l_{n_\epsilon}+j}^\epsilon(\theta_j, \xi_j).$$

Let $\epsilon n_\epsilon \rightarrow 0$. Now, using the uniform weak continuity of the P_k^ϵ , the uniform continuity of the G_k^ϵ , the tightness (3.4), and the fact that

$$\sup_{0 \leq j \leq n_\epsilon} |\theta_{l_{n_\epsilon}+j}^\epsilon - \theta_{l_{n_\epsilon}}^\epsilon| \rightarrow 0,$$

we can work backwards in the above equation, successively concentrating the measure of $\theta_{l_{n_\epsilon}+j}^\epsilon$ at $\theta_{l_{n_\epsilon}}^\epsilon$ and ultimately yielding the representation

$$\int \cdots \int P_{l_{n_\epsilon}}^\epsilon(\xi_{l_{n_\epsilon}}^\epsilon, d\xi_1 | \theta_{l_{n_\epsilon}}^\epsilon) \cdots P_{l_{n_\epsilon}+j-1}^\epsilon(\xi_{j-1}, d\xi_j | \theta_{l_{n_\epsilon}}^\epsilon) G_{l_{n_\epsilon}+j}^\epsilon(\theta_{l_{n_\epsilon}}^\epsilon, \xi_j) + \rho_{l_{n_\epsilon}}^\epsilon(\epsilon, j).$$

The error term satisfies $|\rho_{l_{n_\epsilon}}^\epsilon(\epsilon, j)| \leq \rho(\epsilon, n_\epsilon)$, where $\rho(\epsilon, n_\epsilon)$ depends on the moduli of continuity and

$$(A2.2) \quad \rho(\epsilon, n_\epsilon) \rightarrow 0$$

as $\epsilon \rightarrow 0$ for each constant $n_\epsilon = m$. Consequently, there are $n_\epsilon \rightarrow \infty$ such that (A2.2) holds for this sequence. The above discussion and the proof of Theorem 3.1 imply that the conclusions of Theorem 3.1 will hold under the additional condition that there is a function $g(\cdot)$ (which must be continuous by the above arguments) such that

$$(A2.3) \quad \frac{1}{n_\epsilon} \sum_{j=l_{n_\epsilon}}^{l_{n_\epsilon}+n_\epsilon-1} E_{l_{n_\epsilon}}^\epsilon G_j^\epsilon(\theta, \xi_j(\theta)) \rightarrow g(\theta)$$

in mean for each θ , as $l \rightarrow \infty$, $\epsilon \rightarrow 0$, and $n_\epsilon \rightarrow \infty$.

An advantage of this averaging approach is that it can be used for grouping terms when the dependence of G_j^ϵ on ϵ, j does not vanish for large j and small ϵ .

Appendix 3. Arbitrary updatings within a regeneration period. This appendix illustrates the possibilities when updates are made after “partial” observations. It is intended to be suggestive and is a little vague. To ensure that the “partial” observations fit together properly, additional conditions are needed. Recall the example of §7, where we updated after each N departures. Owing to the regeneration structure of the problem, one could have updated at the end of each regeneration period if the conditions of §3.4 held. These two approaches yield two different time scales in which to get the limit results. The $g(\cdot)$ functions would be different in the two cases but are related by the constant, which is the mean length of the renewal period. The results are equivalent since the two ODEs have the same asymptotic behavior. As seen in §3 and in the examples, there is no need in general to update at the end of regeneration periods. Indeed, even if the problem admits of a regeneration model, for general problems the intervals might be excessively long. If the problem has the “atomic increment” property of §3.3, then the regenerative structure does allow a rather arbitrary method of updating, within the intervals. By a regeneration process, we mean that for each fixed θ the process is regenerative and that for the physical process with the varying θ the conditional distribution of functionals of the

intervals $n, n+1, \dots$ given the past depends only on the parameter value at the start of the n th interval. We will work within the regeneration setup but wish to update at arbitrary intervals (random times). This falls easily and naturally into our framework, as will now be shown. Only a brief outline will be given.

The basic algorithm is still (3.1). The updating times within the regeneration intervals can be chosen rather arbitrarily, subject to the mild conditions below. But we always update at the end of each regeneration interval. This last condition is not necessary but does simplify the discussion. Otherwise the groupings of the terms would be more involved. Let N_n^ϵ denote the number of updates in the n th regeneration interval, $n = 1, 2, \dots$. Define $M_0^\epsilon = 0$ and $M_n^\epsilon = \sum_{i=1}^{n-1} N_i^\epsilon$, $n \geq 1$. We now state the basic redefinitions and assumptions. They are essentially copies of those of Theorem 3.1. But since the estimation process begins anew at the start of each regeneration interval, the assumptions concern what happens within the intervals.

Let $N_n^\epsilon < \infty$ w.p.1 for all $\epsilon > 0, n < \infty$. Let $Y_{n,j}^\epsilon, j = 0, \dots, N_n^\epsilon - 1$ denote the observations in the n th regeneration interval. We update after each observation. Hence there will be N_n^ϵ updates in the n th interval. Due to the assumption of a regenerative structure, the $\{Y_{m,i}^\epsilon, m \geq n, i \geq 0\}$ are conditionally independent of $\{Y_{m,i}^\epsilon, m < n, i \geq 0\}$ given $\theta_{M_n^\epsilon}^\epsilon$, the parameter value at the start of the n th interval. For $j \geq N_n^\epsilon$, set $Y_{n,j}^\epsilon = 0$. For each $\epsilon > 0, n \geq 1$, let $\mathcal{B}_{n,j}^\epsilon$ be a nondecreasing sequence of sigma-algebras measuring at least $\{\theta_{M_n^\epsilon}^\epsilon, Y_{n,i}^\epsilon, i < j\}$, with $E_{n,j}^\epsilon$ denoting the associated conditional expectation. Assume

$$(A3.1a) \quad \left\{ \sum_{j=0}^K |Y_{n,j}^\epsilon|; n, \epsilon \right\} \text{ is uniformly integrable for each } K,$$

$$(A3.1b) \quad E \sum_{N_n^\epsilon \wedge K}^{N_n^\epsilon} |Y_{n,j}^\epsilon| \rightarrow 0 \text{ as } K \rightarrow \infty.$$

Remark on (A3.1) for the example of §7. Return to the physical problem of §7. Suppose that there is an integer M such that we update at least after each new M departures but otherwise use the updating model of this section. Let R_n^ϵ denote the number of customers in the n th regeneration interval, and set $Q_n^\epsilon = \sum_{i=1}^{n-1} R_i^\epsilon$. We have

$$\sum_{i=0}^{N_n^\epsilon-1} |Y_{n,i}^\epsilon| \leq \sum_{i=1}^{R_n^\epsilon} \sum_{l=1}^i |Z_{Q_n^\epsilon+l}^\epsilon|.$$

Condition (A3.1a) holds if $\{Z_l^\epsilon, \epsilon, l\}$ is uniformly integrable. Condition (A3.1b) holds if

$$\lim_{K \rightarrow \infty} E \sum_{i=K \wedge R_n^\epsilon}^{R_n^\epsilon} \sum_{l=1}^i |Z_{Q_n^\epsilon+l}^\epsilon| = 0,$$

where the limit is taken on uniformly in (n, ϵ) .

The SA algorithm and interpolation. Now define $\theta_{n,j}^\epsilon = \theta_{M_n^\epsilon+j}^\epsilon$. The algorithm within the n th interval is

$$(A3.2) \quad \theta_{n,j+1}^\epsilon = \theta_{n,j}^\epsilon + \epsilon Y_{n,j}^\epsilon, \quad j < N_n^\epsilon.$$

Define the interpolated process $\theta^\epsilon(\cdot)$ by $\theta^\epsilon(\cdot) = \theta_{M_n^\epsilon}^\epsilon$ on the interval $[\epsilon n, \epsilon n + \epsilon)$, $\theta^\epsilon(t) = \theta_0^\epsilon$, $t \leq 0$. Thus, we update the parameter at arbitrary times but define the interpolation $\theta^\epsilon(\cdot)$ by the values of the parameter at the end of the regeneration intervals only. This makes the scaling easier, allows a nicer representation of the limit ODE, and yields the desired limit points of the algorithm. Assume that

$$(A3.3) \quad \{\theta_{M_n^\epsilon}^\epsilon; \epsilon > 0, n < \infty\} \text{ is tight.}$$

Analogous to the situation in Theorem 3.1, suppose that there are random variables $\{\xi_{n,j}^\epsilon; \epsilon > 0, j < \infty\}$ and measurable $G_{n,j}^\epsilon(\cdot)$ such that for $j < N_n^\epsilon$

$$E_{n,j}^\epsilon Y_{n,j}^\epsilon = G_{n,j}^\epsilon(\theta_{n,j}^\epsilon, \xi_{n,j}^\epsilon).$$

The values of $\xi_{n,j}^\epsilon$ for $j \geq N_n^\epsilon$ are irrelevant, and one can use any convenient one. Let $G_{n,j}^\epsilon(\cdot)$ be continuous, uniformly in ϵ, n, j and on each compact (θ, ξ) set. Assume

$$(A3.4) \quad \{\xi_{n,j}^\epsilon; \epsilon > 0, n < \infty, j < \infty\} \text{ is tight.}$$

Suppose that there are transition functions $P_{n,j}^\epsilon(\cdot)$ such that $P_{n,j}^\epsilon(\cdot, A|\cdot)$ is measurable for each Borel set A and that

$$(A3.5) \quad P_{n,j}^\epsilon(\xi_{n,j}^\epsilon, \xi_{n,j+1}^\epsilon \in \cdot | \theta_{n,j}^\epsilon) = P\{\xi_{n,j+1}^\epsilon \in \cdot | \theta_{M_n^\epsilon}^\epsilon, Y_{n,k}^\epsilon, k \leq j\}.$$

Let $P_{n,j}^\epsilon(\xi, d\tilde{\xi}|\theta)$ be weakly continuous in (θ, ξ) , uniformly in ϵ, n, j and in each compact (θ, ξ) set. Now for each n, ϵ and θ , $P_{n,j}^\epsilon(\cdot|\theta)$, $j \geq 0$, defines a nonhomogeneous "fixed θ " Markov chain. Let $\{\xi_{n,j}^\epsilon(\theta), j = 0, 1, \dots\}$ denote the random variables of this chain.

Assume that there are continuous functions $g_n^\epsilon(\cdot)$ such that

$$(A3.6) \quad g_n^\epsilon(\theta) = E \sum_{j=0}^{N_n^\epsilon-1} G_{n,j}^\epsilon(\theta, \xi_{n,j}^\epsilon(\theta)).$$

Define $\hat{\theta}_n^\epsilon = \theta_{M_n^\epsilon}^\epsilon$. Let there be a continuous function $g(\cdot)$ such that for each $\delta > 0$

$$(A3.7) \quad \lim_{\epsilon} \limsup_n P\{|g_{n+1}^\epsilon(\hat{\theta}_n^\epsilon) - g(\hat{\theta}_n^\epsilon)| \geq \delta\} = 0.$$

THEOREM A3.1. *Assume the conditions of this section. Then the conclusions of Theorem 3.1 continue to hold for $\hat{\theta}^\epsilon(\cdot)$ and $\hat{\theta}^\epsilon(\epsilon q_\epsilon + \cdot)$ and similarly for Theorems 3.2, 3.3, 4.1, and 5.1.*

Proof. Only a few basic remarks will be made. The assumption (A3.1) allows the set of observations in an interval to be truncated to and well approximated by some finite number K and guarantees uniform integrability of the set of those truncations. By (A3.1) and following the scheme in Theorem 3.1, for $n > 0$ we can write

$$(A3.8) \quad \hat{\theta}_n^\epsilon = \theta_0^\epsilon + \epsilon \sum_{i=1}^n E_{i,0}^\epsilon \sum_{j=0}^{N_i^\epsilon-1} G_{i,j}^\epsilon(\theta_{i,j}^\epsilon, \xi_{i,j}^\epsilon) + \rho_n^\epsilon,$$

where, for each t , $\sup_{\epsilon, n: \epsilon n \leq t} E|\rho_n^\epsilon| \rightarrow 0$ as $\epsilon \rightarrow 0$. Indeed, (A3.1) implies the tightness of $\{\hat{\theta}^\epsilon(\cdot), \hat{\theta}^\epsilon(\epsilon q_\epsilon + \cdot), \epsilon > 0\}$ and the fact that any weak limit will have Lipschitz

continuous paths w.p.1. The assumed uniform θ continuity of $G_{n,j}^\epsilon(\cdot)$, (A3.1), and the fact that

$$\sup_{j < N_i^\epsilon} |\theta_{i,j}^\epsilon - \hat{\theta}_{i-1}^\epsilon| \rightarrow 0$$

(in probability, uniform in i) as $\epsilon \rightarrow 0$ imply that we can replace the $\theta_{i,j}^\epsilon$ in (A3.8) with $\hat{\theta}_{i-1}^\epsilon$ without affecting the limit.

The only remaining problem concerns the fact that the distribution of the $\xi_{i,j}^\epsilon$ in (A3.8) depends on all the $\theta_{i,k}^\epsilon, k < j$. But the representation (A3.5), condition (A3.1), and the asserted uniform continuity of the $P_{i,j}^\epsilon, G_{i,j}^\epsilon$ can be used to show that $\xi_{i,j}^\epsilon$ can be replaced with $\xi_{i,j}^\epsilon(\hat{\theta}_{i-1}^\epsilon)$ without changing the limits. Finally, (A3.6) and (A3.7) are used to complete the proof of the analogue of Theorem 3.1. The analogues of the other theorems will then follow. $\square$

Appendix 4. Distributed/asynchronous updating: A network example. We will discuss a useful canonical form for a SA procedure that operates in a decentralized way and where different components of the iterate might be updated at different (random) times. Some components might be updated much more frequently than others. This is typical of a growing number of applications. One example is given below. Owing to this asynchronous behavior between the components, one needs to work with interpolated processes in an appropriate *real time* scale. Heretofore, the interpolations were based on the iterate number. But now, due to the possibly different times and frequencies of updating of the different components, one needs to use a common time scale for all the components, and this will be an appropriate “real” time scale. The general idea of the proof is just that of Theorem 3.1. The main added feature concerns the difference in the time scaling of the interpolations. Working directly with the iterates can lead to a notational nightmare. We avoid the need to deal directly with the possibly different and random interpolation intervals by using appropriate rescaling. This puts the problem into a framework where the previous results can be directly applied. The result is a simplification and extension of the results in [30, 40], where the ideas of time scaling first appeared. The central idea of the rescaling is easier to see if we start with a centralized and synchronized updating and interpolate in real time. This will be done in the subsection below. The general result for the decentralized problem is in §A4.2.

A4.1. A synchronized updating: Real time scale. Assume the conditions of Theorem 3.1. The algorithm is

$$\theta_{n+1}^\epsilon = \theta_n^\epsilon + \epsilon Y_n^\epsilon.$$

Then Theorem 3.1 holds for the interpolations $\{\theta^\epsilon(\cdot), \theta^\epsilon(\epsilon q_\epsilon + \cdot)\}$.

Now let us rewrite the interpolation in real time. Let $\delta\tau_n^\epsilon$ denote the time interval between the n th and $(n+1)$ st updating. Let $\mathcal{B}_n^\epsilon$ be a nondecreasing sequence of sigma-algebras such that $\mathcal{B}_n^\epsilon$ measures at least $\{\theta_0^\epsilon, Y_i^\epsilon, \delta\tau_i^\epsilon, i < n\}$. Let E_n^ϵ denote the associated conditional expectation. We keep the same framework as in §3. Suppose that there is a Markov chain ξ_n^ϵ and a continuous and strictly positive function $u(\cdot)$ such that $E_n^\epsilon \delta\tau_n^\epsilon = u(\theta_n^\epsilon, \xi_n^\epsilon)$. Assume the tightness condition (3.4), the uniform integrability condition (3.2), and

$$(A4.1) \quad \{\delta\tau_n^\epsilon, n < \infty\} \text{ is uniformly integrable.}$$

Also suppose that there is a continuous and bounded function $\hat{u}(\cdot)$ such that

$$(A4.2) \quad \frac{1}{n} \sum_{i=0}^{n-1} Eu(\theta, \xi_i(\theta)) \rightarrow \hat{u}(\theta)$$

for each θ and initial condition.

Define

$$\tau_n^\epsilon = \epsilon \sum_{i=0}^{n-1} \delta \tau_i^\epsilon,$$

$$N^\epsilon(t) = \epsilon [\text{number of updatings by time } t/\epsilon].$$

Let $\tau^\epsilon(\cdot)$ be the interpolation of $\{\tau_n^\epsilon, n < \infty\}$ defined by $\tau^\epsilon(t) = \tau_n^\epsilon$ on $[\epsilon n, \epsilon(n+1))$. Note that $\tau^\epsilon(\cdot)$ is the inverse of $N^\epsilon(\cdot)$ in the sense that $N^\epsilon(\tau^\epsilon(t)) = n\epsilon$ for $t \in [n\epsilon, (n+1)\epsilon)$ and

$$(A4.3) \quad \tau^\epsilon(t) = \inf\{s : N^\epsilon(s) \geq t\}.$$

Define $\hat{\theta}^\epsilon(t) = \theta^\epsilon(N^\epsilon(t))$. This is the interpolation in the real time scale, not the iterate time scale. The weak convergence and characterization of the ODE for the $\hat{\theta}^\epsilon(\cdot)$ are now easily done. In all cases we suppose that the original sequence indexed by ϵ converges weakly. Otherwise take appropriate subsequences. By the above conditions and the proof of Theorem 3.1, $(\tau^\epsilon(\cdot), \hat{\theta}^\epsilon(\cdot), N^\epsilon(\cdot), \theta^\epsilon(\cdot)) \Rightarrow (\tau(\cdot), \hat{\theta}(\cdot), N(\cdot), \theta(\cdot))$, where

$$(A4.4) \quad \hat{\theta}(t) = \theta(N(t)) \quad \text{and} \quad \tau(t) = \int_0^t \hat{u}(\theta(s)) ds.$$

From the positivity of $\hat{u}(\theta)$ and the “inverse” definitions of $N^\epsilon(\cdot)$ and $\tau^\epsilon(\cdot)$, it follows that $N^\epsilon(\cdot) \Rightarrow N(\cdot)$, where $N(\tau(t)) = t$. Taking derivatives, we get $\dot{N}(\tau(t))\dot{\tau}(t) = 1$. Call $s = \tau(t)$. Then using (A4.4), the slope of $N(s)$ is $\dot{N}(s) = 1/\hat{u}(\theta(\tau^{-1}(s))) = 1/\hat{u}(\hat{\theta}(s))$. Therefore

$$(A4.5) \quad N(t) = \int_0^t \frac{ds}{\hat{u}(\hat{\theta}(s))}.$$

By Theorem 3.1, $\theta(\cdot)$ satisfies $\dot{\theta} = g(\theta)$. Recall that

$$(A4.6) \quad \hat{\theta}^\epsilon(\cdot) = \theta^\epsilon(N^\epsilon(\cdot)) \Rightarrow \theta(N(\cdot)) \equiv \hat{\theta}(\cdot).$$

Thus, using the fact that $N(\tau(t)) = t$, we can write

$$(A4.7) \quad \dot{\hat{\theta}}(t) = [\dot{\theta}(N(t))] \dot{N}(t) = g(\theta(N(t))) / \hat{u}(\hat{\theta}(t)) = g(\hat{\theta}(t)) / \hat{u}(\hat{\theta}(t)).$$

Thus, the proof is just Theorem 3.1 plus a time change argument. The purpose of the time change argument is to avoid dealing with random interpolation intervals and the interaction of the Y_n^ϵ and the $\delta \tau_n^\epsilon$. It exploits the convergence of both the “time” processes and of the original interpolation $\theta^\epsilon(\cdot)$.

Remark. The above argument is for processes that start at time zero with limits defined on the interval $[0, \infty)$. Suppose that we wish to get the limit on $(-\infty, \infty)$ of

$\hat{\theta}^\epsilon(T^\epsilon + \cdot)$, where T^ϵ is a sequence of real numbers tending to infinity. The T^ϵ is simply the replacement for the ϵq_ϵ in Theorem 3.1. Then the analysis is the same as above, except that the initial condition of the interpolation is

$$\hat{\theta}^\epsilon(T^\epsilon) = \theta^\epsilon(N^\epsilon(T^\epsilon)) = \theta_{N^\epsilon(T^\epsilon)/\epsilon}^\epsilon,$$

the values of the parameter at increasingly large iterate numbers, and the uniform integrability and tightness conditions must reflect this change. In particular, we need tightness of

$$(A4.8) \quad \{\xi_{N^\epsilon(t)/\epsilon+n}^\epsilon, \theta_{N^\epsilon(t)/\epsilon+n}^\epsilon; \epsilon, n, t\}$$

and uniform integrability of

$$(A4.9) \quad \{Y_{N^\epsilon(t)/\epsilon+n}^\epsilon, \delta\tau_{N^\epsilon(t)/\epsilon+n}^\epsilon; \epsilon, n, t\}.$$

Since all sorts of dependencies among the two sequences Y_n^ϵ and $\delta\tau_n^\epsilon$ can be constructed, little can be said without further assumptions. But a casual examination of some simple cases suggests that (A4.8) and (A4.9) are not very restrictive.

A distributed and decentralized network model. We work with one canonical model in order to illustrate some of the possibilities and minimize notation. To simplify the notation, some of the conditions will be less general than can be handled by the introduced technique. The basic work is in setting up the notation for the various time scales. Basically the general method uses the idea of the above subsection separately on different parts of the problem, as will now be seen.

Let $\theta = (\theta_1, \dots, \theta_K)$, where the θ_α are the scalar components of θ . Consider a system with K controllers, each of which is responsible for the updating of one component. We wish to minimize a function $F(\cdot)$ which takes the form $F(\theta) = \sum_{\beta=1}^K F^\beta(\theta)$ for real-valued and continuously differentiable $F^\beta(\cdot)$. Let $F_\alpha^\beta(\theta) = \partial F^\beta(\theta) / \partial \theta_\alpha$. In our model, for each α subsystem β produces a sequence of estimates $Y_{\alpha,n}^{\beta,\epsilon}$, $n = 0, \dots$, which it sends to node α for help in estimating F_α^β at whatever the current value of θ is. It also sends the current values of its own component θ_β .

Example. An important class of examples that provides a guide to the development are the problems of optimal routing in queueing networks. Let the network have K nodes, with θ the K vector of routing parameters, where θ_α is the component associated with the α th node. Let $F^\beta(\theta)$ denote the stationary average queue length at node β under parameter value θ . We wish to minimize the stationary average number of customers in the network $F(\theta) = \sum F^\beta(\theta)$. The problem arises in control of telecommunication networks and has been treated in [42, 40]. The controller at node α updates the component θ_α of θ , and it does so based on both its own observations and relevant data sent from other nodes. In one useful approach, called the *surrogate estimation method* in the above references, each node β estimates the sensitivity of the mean length of its own queue to variations in *external* inputs to that node. Then one uses the mean systems flow equations to get acceptable estimates of the $F_\alpha^\beta(\theta)$. These estimates are transmitted to node α for use in estimating the derivative of $F(\theta)$ with respect to θ_α at the current value of θ and then updating the value of θ_α . After each transmission, new estimates are taken and the process is repeated. The method gave good results in simulations.

The times required for the estimation intervals can depend heavily and randomly on the node. They might be functions of the number of service completions or simply

deterministic time intervals. The nodes would transmit their estimates in an asynchronous way. Thus the SA is both decentralized and unsynchronized. In general, θ_α would be a vector of routing probabilities. For simplicity of notation, we shall consider only scalar components. The extensions to the vector case are straightforward. In a typical application of SA, each time a new estimate of $F_\alpha^\beta(\theta)$ (at the current value of θ) is received at node α , that estimate is multiplied by a step-size parameter and subtracted from the current value of state component θ_α . This “additive” structure allows us to represent the algorithm in a useful decomposed way by writing the current value of the component θ_α as the sum of the initial value plus K terms. The β th such term is the product of an appropriate step-size times the sum of the past transmissions from node β to node α of the estimates of $F_\alpha^\beta(\theta)$ at whatever the operating values of the parameter were when the estimates were made. In the development below, this decomposition is formalized and provides a useful simplification.

We shall now return to our general model. The time for transmission of information can have bounded delays, and these delays cause no problems in the analysis. But only to simplify notation, we work under the assumption that there are no delays and that the parameters are updated as soon as new information is available. The reader can fill in the few additional details for the delayed case. We are reluctant to try a very general development since the entire field of decentralized/asynchronous optimization is in its infancy, and one expects many new models and methods for estimation to appear in the next few years. But the methods employed would be fundamental to any extensions.

Notation. Let $\delta\tau_{\alpha,n}^{\beta,\epsilon}$ denote the interval between the n th and $(n+1)$ st transmissions from β to α . Define

$$\tau_{\alpha,n}^{\beta,\epsilon} = \epsilon \sum_{i=0}^{n-1} \delta\tau_{\alpha,i}^{\beta,\epsilon},$$

ϵ times the real time required by the first n transmissions from β to α . Define

$$N_\alpha^{\beta,\epsilon}(t) = \epsilon [\text{number of transmissions from } \beta \text{ to } \alpha \text{ to reach real time } t/\epsilon].$$

Let $\tau_{\alpha,n}^{\beta,\epsilon}(t)$ be the interpolation of the $\tau_{\alpha,n}^{\beta,\epsilon}$ with interpolation intervals ϵ and initial condition zero. Analogously to the situation in the last subsection, $N_\alpha^{\beta,\epsilon}(\cdot)$ and $\tau_{\alpha,n}^{\beta,\epsilon}(\cdot)$ are inverses of one another.

The SA algorithm. The notation is a little complex but very natural. It enables us to carry over the results of Theorem 3.1 to a much more complex situation via several time-change arguments and thus saves a great deal of work over a direct analysis. Let $\hat{\theta}^\epsilon(\cdot) = \{\hat{\theta}_i^\epsilon(\cdot), i \leq K\}$ denote the interpolation in the real time (times ϵ) scale. As mentioned in the discussion of the example, it is convenient to separate $\hat{\theta}_\alpha^\epsilon(\cdot)$ into components which come from the same node. This suggests the following decomposed representation for the SA algorithm. For each α, β , let $c_\alpha^\beta(\cdot)$ be a continuous and bounded real-valued function and define the sequence $\theta_{\alpha,n}^{\beta,\epsilon}$ by

$$(A4.10) \quad \theta_{\alpha,n+1}^{\beta,\epsilon} = \theta_{\alpha,n}^{\beta,\epsilon} + \epsilon c_\alpha^\beta(\hat{\theta}^\epsilon(\tau_{\alpha,n}^{\beta,\epsilon})) Y_{\alpha,n}^{\beta,\epsilon}.$$

The role of the $c_\alpha^\beta(\cdot)$ functions is to partially compensate for the fact that the frequency of the intervals between updates might depend on θ, α, β , and will be further commented upon at the end of the section. In many cases, we would use $c_\alpha^\beta(\theta) \equiv 1$. Note

that by the definitions $\hat{\theta}^\epsilon(\tau_{\alpha,n}^{\beta,\epsilon})$ is the state value at the time of the n th transmission from node β to node α . Indeed, we can write

$$(A4.11) \quad \begin{aligned} \hat{\theta}_\alpha^{\beta,\epsilon}(t) &= \theta_{\alpha, N_{\alpha,n}^{\beta,\epsilon}(t)/\epsilon}^{\beta,\epsilon} = \theta_\alpha^{\beta,\epsilon}(N_\alpha^{\beta,\epsilon}(t)), \\ \hat{\theta}_\alpha^{\beta,\epsilon}(\tau_{\alpha,n}^{\beta,\epsilon}(t)) &= \theta_\alpha^{\beta,\epsilon}(t), \text{ where } \theta_\alpha^{\beta,\epsilon}(t) = \theta_{\alpha,n}^{\beta,\epsilon} \text{ for } t \in [n\epsilon, (n+1)\epsilon). \end{aligned}$$

We can now define the actual interpolated iterate in the appropriate real time scale in terms of the components as

$$(A4.12) \quad \hat{\theta}_\alpha^\epsilon(t) = \hat{\theta}_\alpha^\epsilon(0) + \sum_{\beta=1}^K \hat{\theta}_\alpha^{\beta,\epsilon}(t), \quad \hat{\theta}_\alpha^{\beta,\epsilon}(0) = 0.$$

It will be shown that the proofs are just adaptations of the argument in the last subsection to the vector case. It will be seen from the argument that all sorts of groupings and variations can be added to the format.

Assumptions. Let $\mathcal{B}_t^\epsilon$ be a nondecreasing sequence of sigma-algebras which measure at least the initial conditions and all the data transmitted by all the nodes up to real time t . Let $E_{\alpha,n}^{\beta,\epsilon}$ equal the expectation conditioned on $\mathcal{B}_{\tau_{\alpha,n}^{\beta,\epsilon}/\epsilon}^{\beta,\epsilon}$. Thus, $E_{\alpha,n}^{\beta,\epsilon}$ can be interpreted as the expectation conditioned on the information which is available at the time $\tau_{\alpha,n}^{\beta,\epsilon}$ of the n th transmission from β to α . We next give the conditions on the interupdate intervals. We suppose that for each α, β, ϵ there is a Markov chain $\{\xi_{\alpha,n}^{\beta,\epsilon}, n < \infty\}$ whose transition functions satisfy the obvious analogue of (3.5) and continuous functions $u_\alpha^\beta(\cdot)$ such that

$$(A4.13) \quad \{\delta\tau_{\alpha,n}^{\beta,\epsilon}; \alpha, \beta, \epsilon, n\} \text{ is uniformly integrable,}$$

$$(A4.14) \quad E_{\alpha,n}^{\beta,\epsilon} \delta\tau_{\alpha,n}^{\beta,\epsilon} = u_\alpha^\beta(\hat{\theta}_\alpha^\epsilon(\tau_{\alpha,n}^{\beta,\epsilon}), \xi_{\alpha,n}^{\beta,\epsilon}).$$

Let there be fixed θ Markov chains $\{\xi_{\alpha,n}^\beta(\theta)\}$ with transition probabilities satisfying the analogues of (3.6)–(3.8). Let there be continuous functions $\hat{u}_\alpha^\beta(\cdot)$ such that for each initial condition and each θ

$$(A4.15) \quad \lim_N \frac{1}{N} \sum_{i=0}^{N-1} E u_\alpha^\beta(\theta, \xi_{\alpha,n}^\beta(\theta)) = \hat{u}_\alpha^\beta(\theta).$$

We also need that

$$(A4.16) \quad \inf_{\theta, \xi} u_\alpha^\beta(\theta, \xi) > 0.$$

Let there be uniform integrability of

$$(A4.17) \quad \{Y_{\alpha,n}^{\beta,\epsilon}, \alpha, \beta, \epsilon, n\}$$

and tightness of

$$(A4.18) \quad \{\xi_{\alpha,n}^{\beta,\epsilon}, \theta_{\alpha,n}^{\beta,\epsilon}; \alpha, \beta, \epsilon, n\}.$$

Define $G_\alpha^\beta(\cdot)$ in the usual way:

$$E_{\alpha,n}^{\beta,\epsilon} Y_{\alpha,n}^{\beta,\epsilon} = G_\alpha^\beta(\hat{\theta}_\alpha^\epsilon(\tau_{\alpha,n}^{\beta,\epsilon}), \xi_{\alpha,n}^{\beta,\epsilon}) + \text{small error},$$

where the small error goes to zero in the mean as $\epsilon \rightarrow 0$ and $n \rightarrow \infty$. Suppose that there are continuous functions $g_\alpha^\beta(\theta)$ such that for each initial condition and each θ

$$(A4.19) \quad \lim_N \frac{1}{N} \sum_{n=1}^N EG_\alpha^\beta(\theta, \xi_{\alpha,n}^\beta(\theta)) = g_\alpha^\beta(\theta).$$

The above conditions are for starting at time zero. If we wish to work with $\hat{\theta}^\epsilon(T^\epsilon + \cdot)$ as in the first subsection, then we need to shift the indices analogously to what was done there.

Remark. In the case of the example mentioned above, the assumptions on the interupdate intervals are obviously satisfied if the nodes compute the estimators at constant intervals but also in many cases where (for example) either a fixed number of service completions or perhaps a “local” regenerative approach is used for the local estimation. Note that the chains are “local” in the sense that they can depend on the pair α, β . Thus, we work with each pair separately, which can give a simpler chain than what would appear if we treated all the pairs simultaneously. Equation (A4.16) is used to guarantee that $\{N_\alpha^{\beta,\epsilon}(\cdot), \epsilon > 0\}$ is tight and has continuous limits.

THEOREM A4.1. *Every subsequence of $\hat{\theta}(\cdot)$ has a further subsequence that converges weakly to a solution of the ODE:*

$$\dot{\hat{\theta}}_\alpha = \sum_{\beta=1}^K \frac{c_\alpha^\beta(\hat{\theta})}{\hat{u}_\alpha^\beta(\hat{\theta})} g_\alpha^\beta(\hat{\theta}).$$

Comments. Now we see the use of the $c_\alpha^\beta(\cdot)$ functions as a way of dealing with the variable $\hat{u}_\alpha^\beta(\theta)$. For the example, the $g_\alpha^\beta(\theta)$ are supposed to be approximations to the $F_\alpha^\beta(\theta)$. The proof of the theorem uses the ideas of the previous subsection. The $\{\hat{\theta}_\alpha^{\beta,\epsilon}(\cdot), \tau_\alpha^{\beta,\epsilon}(\cdot)\}$ is tight, and all weak limits have Lipschitz-continuous paths w.p.1. Also, $\{N_\alpha^{\beta,\epsilon}(\cdot)\}$ is tight, and all limits have Lipschitz-continuous paths. Thus, $\{\hat{\theta}_\alpha^{\beta,\epsilon}(\cdot), \hat{\theta}^\epsilon(\cdot)\}$ is tight and has Lipschitz-continuous limits. Let $(\theta_\alpha^\beta(\cdot), \theta(\cdot), \hat{\theta}(\cdot), \tau_\alpha^\beta(\cdot), N_\alpha^\beta(\cdot))$ denote the limit processes. We have $\tau_\alpha^{\beta,\epsilon}(\cdot) \Rightarrow \tau_\alpha^\beta(\cdot)$, where

$$\tau_\alpha^\beta(t) = \int_0^t \hat{u}_\alpha^\beta(\hat{\theta}(\tau_\alpha^\beta(s))) ds.$$

In the centralized case of the previous subsection, the argument of the $\hat{u}$ reduces to just $\theta(s)$. Also $N_\alpha^{\beta,\epsilon}(\cdot) \Rightarrow N_\alpha^\beta(\cdot)$, where

$$N_\alpha^\beta(t) = \int_0^t \frac{ds}{\hat{u}_\alpha^\beta(\hat{\theta}(s))}.$$

The form of the algorithm (A4.10), the weak convergence of $(\hat{\theta}^\epsilon(\cdot), \tau_\alpha^{\beta,\epsilon}(\cdot))$ to $(\hat{\theta}(\cdot), \tau_\alpha^\beta(\cdot))$, and Theorem 3.1 yield that

$$\dot{\theta}_\alpha^\beta(t) = c_\alpha^\beta(\hat{\theta}(\tau_\alpha^\beta(t))) g_\alpha^\beta(\hat{\theta}(\tau_\alpha^\beta(t))).$$

The theorem follows by writing the expression for $\dot{\theta}_\alpha^\beta(t)$ and using the fact that $N_\alpha^\beta(\tau_\alpha^\beta(t)) = t$.

Remark. Note the great advantage in using the rescaling idea. It allows us to separate the intervals from the values of the updates in the analysis and permits a result under quite weak conditions with minimal new work. It is a technique of considerable utility. The analogues of Theorems 3.2, 4.1, and 5.1 also hold.

REFERENCES

- [1] M. BENAÏM, *A dynamical systems approach to stochastic approximations*, Univ. of California, Berkeley, CA, 1993, preprint.
- [2] A. BENVENISTE, M. METIVIER, AND P. PRIOURET, *Adaptive Algorithms and Stochastic Approximation*, Springer-Verlag, New York, Berlin, 1990.
- [3] P. BILLINGSLEY, *Convergence of Probability Measures*, John Wiley, New York, 1968.
- [4] P. BRÉMAUD AND F. J. VÁZQUEZ-ABAD, *On the pathwise computation of derivatives with respect to the rate of a point process: The phantom RPA method*, Queueing Systems Theory Appl., 10 (1992), pp. 249–270.
- [5] F. CAMPILLO AND E. PARDOUX, *Numerical methods in ergodic optimal stochastic control*, in Applied Stochastic Analysis, I. Karatzas and D. Ocone, eds., Springer-Verlag, Berlin, 1991, pp. 59–73.
- [6] M. CARAMANIS AND G. LIBEROPOULOS, *Perturbation analysis for the design of flexible manufacturing system flow controllers*, Oper. Res., 40 (1992), pp. 1107–1125.
- [7] E. K. P. CHONG AND P. J. RAMADGE, *Optimization of queues using an infinitesimal perturbation analysis-based stochastic algorithm with general update times*, SIAM J. Control Optim., 31 (1993), pp. 698–732.
- [8] M. H. A. DAVIS, *Markov Models and Optimization*, Chapman-Hall, New York, London, 1993.
- [9] P. DUPUIS AND H. J. KUSHNER, *Stochastic approximation and large deviations: Upper bounds and w.p.1 convergence*, SIAM J. Control Optim., 27 (1989), pp. 1108–1135.
- [10] S. N. ETHIER AND T. G. KURTZ, *Markov Processes: Characterization and Convergence*, John Wiley, New York, 1986.
- [11] M. C. FU, *Convergence of the stochastic approximation algorithm for the GI/G/1 queue using infinitesimal perturbation analysis*, J. Optim. Theory Appl., 65 (1990), pp. 149–160.
- [12] I. I. GIHMAN AND A. V. SKOROHOD, *Stochastic Differential Equations*, Springer-Verlag, Berlin, 1972.
- [13] P. GLASSERMAN, *Gradient Estimation via Perturbation Analysis*, Kluwer, Amsterdam, 1990.
- [14] ———, *Filtered Monte Carlo*, Math. Oper. Res., 18 (1993), pp. 610–634.
- [15] J. GUCKENHEIMER AND P. HOLMES, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag, Berlin, 1983.
- [16] A. HAURIE, P. L'ECUYER, AND CH. VAN DELFT, *Convergence of stochastic approximation coupled with perturbation analysis in a class of manufacturing flow control models*, J. Discrete Event Dynamical Systems, 4 (1994), pp. 87–111.
- [17] Y.-C. HO AND X.-R. CAO, *Performance sensitivity to routing changes in queueing networks and flexible manufacturing systems using perturbation analysis*, IEEE J. on Robotics and Automation, RA-1 (1985), pp. 165–172.
- [18] ———, *Perturbation Analysis of Discrete Event Dynamical Systems*, Kluwer, Boston, 1991.
- [19] Y. C. HO, M. A. EYLER, AND T. T. CHIEN, *A gradient technique for general buffer storage design in a production line*, Internat. J. of Production Research, 17 (1979), pp. 577–580.
- [20] ———, *A new approach to determine parameter sensitivities of transfer lines*, Management Science, 29 (1983), pp. 700–714.
- [21] T. G. KURTZ, *Semigroups of conditional shifts and approximation of Markov processes*, Ann. Probab., 4 (1975), pp. 618–642.
- [22] H. J. KUSHNER, *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic System Theory*, MIT Press, Cambridge, MA, 1984.
- [23] ———, *Robustness and approximation of escape times and large deviations estimates for systems with small noise effects*, SIAM J. Appl. Math., 44 (1984), pp. 160–182.
- [24] ———, *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Systems and Control, Vol. 3, Birkhäuser, Boston, 1990.
- [25] H. J. KUSHNER AND D. S. CLARK, *Stochastic Approximation for Constrained and Unconstrained Systems*, Springer-Verlag, Berlin, New York, 1978.
- [26] H. J. KUSHNER AND J. YANG, *A Monte Carlo method for sensitivity analysis and parametric optimization of nonlinear stochastic systems: The ergodic case*, SIAM J. Control Optim., 30 (1992), pp. 440–464.
- [27] H. J. KUSHNER AND A. SHWARTZ, *An invariant measure approach to the convergence of stochastic approximations with state dependent noise*, SIAM J. Control Optim., 22 (1984), pp. 13–27.
- [28] ———, *Weak convergence and asymptotic properties of adaptive filters with constant gains*, IEEE Trans. Inform. Theory, IT-30 (1984), pp. 177–182.
- [29] H. J. KUSHNER AND J. YANG, *Stochastic approximation with averaging: Optimal asymptotic rates of convergence for general processes*, SIAM J. Control Optim., 31 (1993), pp. 1045–

1062.

- [30] H. J. KUSHNER AND G. YIN, *Stochastic approximation algorithms for parallel and distributed processing*, Stochastics, 22 (1987), pp. 219–250.
- [31] P. L'ECUYER, N. GIROUX, AND P. W. GLYNN, *Stochastic optimization by simulation: Numerical experiments for a simple queue in steady state*, Management Sci., 40 (1994), pp. 1245–1261.
- [32] P. L'ECUYER AND P. W. GLYNN, *Stochastic optimization by simulation: Convergence proofs for the GI/G/1 queue*, Management Sci., 40 (1994), pp. 1562–1578.
- [33] L. LJUNG, *Analysis of recursive stochastic algorithms*, IEEE Trans. Automat. Control, AC-22 (1977), pp. 551–575.
- [34] J. PAN AND C. G. CASSANDRAS, *Flow control of bursty traffic using infinitesimal perturbation analysis*, J. Discrete Event Systems, 4 (1994), pp. 325–358.
- [35] G. CH. PFLUG, *On line optimization of simulated Markov processes*, Math. Oper. Res., 15 (1990), pp. 381–395.
- [36] B. T. POLYAK AND A. B. JUDITSKY, *Acceleration of stochastic approximation by averaging*, SIAM J. Control Optim., 30 (1992), pp. 838–855.
- [37] M. I. REIMAN AND A. WEISS, *Sensitivity analysis for simulation via likelihood ratios*, Oper. Res., 37 (1989), pp. 930–944.
- [38] R. Y. RUBINSTEIN, *Sensitivity analysis and performance extrapolation for computer simulation models*, Oper. Res., 37 (1989), pp. 72–81.
- [39] R. SURI AND B. R. FU, *On using continuous flow lines for performance estimation of discrete production lines*, in Proc. 1991 Winter Simulation Conference, 1991, B.L. Nelson, W. D. Kelton, and G. M. Clark, eds., pp. 968–977.
- [40] F. J. VÁZQUEZ-ABAD, *Stochastic Recursive Algorithms for Optimal Routing in Queueing Networks*, Ph.D. thesis, Brown University, 1989.
- [41] F. J. VÁZQUEZ-ABAD AND H. J. KUSHNER, *Estimation of the derivative of a stationary measure with respect to a control parameter*, J. Appl. Probab., 29 (1992), pp. 343–352.
- [42] ———, *The surrogate estimation approach for sensitivity analysis in queueing networks*, in G. W. Evans, M. Mollaghasemi, E. C. Russel, and W. E. Biles, eds., Proc. Winter Simulation Conference, 1993, pp. 347–355.
- [43] F. J. VÁZQUEZ-ABAD AND L. MASON, *Adaptive decentralized control under non uniqueness of the optimal control*, J. Discrete Event Dyn. Syst., in press.
- [44] H. YAN, G. YIN, AND S. X. C. LOU, *Using stochastic approximation to determine threshold values for control of unreliable manufacturing systems*, J. Optim. Theory Appl., 83 (1994), pp. 511–539.
- [45] H. YAN, X.Y. ZHOU, AND G. YIN, *Optimal numbers of circulating kanbans in a two machine flowshop with uncertainty*, Oper. Res., to appear.
- [46] G. YIN, *On extensions of Polyak's averaging approach to stochastic approximation*, Stochastics, 36 (1992), pp. 245–264.
- [47] ———, *Stochastic approximation via averaging: Polyak's approach revisited*, in Simulation and Optimization: Proceedings of the International Workshops on Computationally Intensive Methods in Simulation and Optimization held at the International Institute for Applied Systems Analysis, Laxenburg, Austria, August 23–25, 1990, Lecture Notes in Economics and Mathematical Systems 374, G. Pflug and U. Dieter, eds., Springer-Verlag, Berlin, 1992, pp. 119–134.