SRC TR 89-76

Stochastic Approximations for Finite-State Markov Chains

By

D.J. Ma, A.M. Makowski
and A. Shwartz

STOCHASTIC APPROXIMATIONS FOR FINITE - STATE MARKOV CHAINS

by

D.-J. Ma[1†], A. M. Makowski[1 †‡] and A. Shwartz[2 ‡]

University of Maryland and Technion

This paper is dedicated to the memory of the late Professor Michel Métivier

ABSTRACT

This paper develops an a.s. convergence theory for a class of projected Stochastic Approximations driven by finite-state Markov chains. The conditions are mild and are given explicitly in terms of the model data, mainly the Lipschitz continuity of the one-step transition probabilities. The approach used here is a version of the ODE method as proposed by Metivier and Priouret. It combines the Kushner-Clark Lemma with properties of the Poisson equation associated with the underlying family of Markov chains.

The class of algorithms studied here was motivated by implementation issues for constrained Markov decision problems, where the policies of interest often depend on quantities not readily available due either to insufficient knowledge of the model parameters or to computational difficulties. This naturally leads to the on-line estimation (or computation) problem investigated here. Several examples from the area of queueing systems are discussed.

Keywords: Stochastic Approximation, Recursive Estimation, Adaptive Control, Markov chains, Poisson equation, Lipschitz continuity.

[1] Electrical Engineering Department and Systems Research Center, University of Maryland, College Park, Maryland 20742, U.S.A. D.-J. Ma is now with Digital Equipment Corporation, MRO1-2/S10, Marlborough, Massachusetts 01752, U.S.A.

[†] The work of these authors was supported partially through NSF Grant ECS-83-51836, partially through a grant from AT&T Bell Laboratories and partially through ONR Grant N00014-84-K-0614.

[2] Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel. The work of this author was supported partially through a Grant from GM Research Laboratories.

[‡] The work of these authors was supported partially through United States—Israel Binational Science Foundation Grant BSF 85-00306.

1. INTRODUCTION

Over the past few years there has been considerable interest in stochastic recursive algorithms. In such schemes, the current output to the algorithm affects the (probabilistic) transition mechanism of a "noise" or "state" process which in turn drives the algorithm in the next iteration. Typically these algorithms produce a sequence of iterates $\{\eta_n, \ n = 0, 1, \ldots\}$ through a recursion

$$\eta_{n+1} = g_n(\eta_n, X_{n+1}) \qquad n = 0, 1, \ldots (1.1)$$

for some Borel measurable mapping $g_n : G \times S \to G$ where G and S are Borel subsets of some Euclidean spaces. The evolution of the S-valued state process $\{X_n, \ n = 0, 1, \ldots\}$ is then characterized by the conditional probability distribution μ_{n+1} of X_{n+1} given $X_0, \eta_0, X_1, \ldots, X_n, \eta_n$.

Of particular interest in this class of algorithms are the schemes first introduced by Robbins and Monro in [15] and known as Stochastic Approximation algorithms. In their simplest form, these algorithms take the form

$$\begin{aligned} \eta_{n+1} &= \eta_n + a_n f(\eta_n, X_{n+1}) \\ \eta_0 &\in \mathbb{R}^p \end{aligned} \qquad n = 0, 1, \ldots (1.2)$$

for some Borel measurable mapping $f : \mathbb{R}^p \times S \to \mathbb{R}^p$, where the sequence of decreasing step sizes $\{a_n, \ n = 0, 1, \ldots\}$ satisfies the standard conditions (2.2).

More recently, it has been necessary to consider projected versions of (1.2), in which case the recursion (1.2) takes the form

$$\begin{aligned} \eta_{n+1} &= \Pi_G \left\{ \eta_n + a_n f(\eta_n, X_{n+1}) \right\} \\ \eta_0 &\in G \end{aligned} \qquad n = 0, 1, \ldots (1.3)$$

where G is a compact convex subset of $\mathbb{R}^p$, Π_G denotes some projection operation on G and f is now a Borel mapping $G \times S \to \mathbb{R}^p$. Usually Π_G is the nearest-point projection on G, but other choices have proved useful [8].

As pointed out earlier, a complete specification of the algorithms (1.2) and (1.3) requires that the one-step transition probability distributions $\{\mu_n, \ n = 1, 2, \ldots\}$ of the state process $\{X_n, \ n = 0, 1, \ldots\}$ be postulated. For instance, the classical Robbins-Monro algorithm [15] corresponds (with $G = \mathbb{R}^p$) to the "i.i.d." case in that

$$P[X_{n+1} \in B | X_0, \eta_0, X_1, \ldots, X_n, \eta_n] = \mu_{\eta_n}(B) \qquad n = 0, 1, \ldots (1.4)$$

for every Borel subset B of S, where $\{\mu_\eta,\ \eta \in G\}$ is a family of probability measures on S. However, motivated by applications, some of which are briefly discussed in Section 3, increasingly complex probabilistic structures have been considered. In particular, Markov dependencies have been found useful in a variety of contexts. This amounts to requiring

$$P[X_{n+1} \in B | X_0, \eta_0, X_1, \ldots, X_n, \eta_n] = \mu_{\eta_n}(X_n; B) \qquad n = 0, 1, \ldots (1.5)$$

for every Borel subset B of S, where $\{\mu_\eta,\ \eta \in G\}$ is a family of one-step probability transition kernels on S. This situation is conveniently referred to as the Markovian case.

The central question in the theory of Stochastic Approximations is concerned with the convergence properties of the iterate sequence $\{\eta_n,\ n = 0, 1, \ldots\}$. In the i.i.d. case (1.2), martingale arguments have been given by Gladyshev [6] to establish a.s. convergence. However in more complex situations, such a direct approach does not work and the so-called ODE method needs to be used. In all its variants, the ODE method proceeds in two separate steps. The first step relies on the Kushner–Clark Lemma in order to identify a deterministic ODE, the stability properties of which determine the limit points of $\{\eta_n,\ n = 0, 1, \ldots\}$. The second step is probabilistic in nature and depends on the algorithm being considered; its purpose is to show that asymptotically (in the mode of convergence of interest) the output sequence to the original algorithm behaves like the solution to the ODE.

In their monograph [8], Kushner and Clark give general conditions for successfully completing this second step. In more structured situations, Kushner has proposed weak convergence methods which require that various tightness properties be established; this leads to convergence in probability of $\{\eta_n,\ n = 0, 1, \ldots\}$. In the Markovian case, Metivier and Priouret [13] establish a.s. convergence by making use of properties of the Poisson equation associated with the transition kernels $\{\mu_\eta,\ \eta \in G\}$ appearing in (1.5). Key to their analysis are properties of Lipschitz continuity (in η) of the solution to this Poisson equation.

Unfortunately, in all these references, the conditions underlying the second step are given in implicit form and are often hard to verify in specific situations. What seems required is a more operational convergence theory where conditions are given *directly* in term of the model data. Although this can probably not be achieved in any great level of generality, it is hoped that such a program can be successfully carried out in structured situations of interest for applications. It is the purpose of this paper to show that a comprehensive convergence theory is available in the Markov case when the state space S is finite and G is a compact convex set of $\mathbb{R}^p$. In that case

(1.5) reduces to

$$P[X_{n+1} = y|X_0, \eta_0, X_1, \ldots, X_n, \eta_n] = p_{X_n y}(\eta_n) \qquad n = 0, 1, \ldots (1.6)$$

for all y in S for some family $\{P(\eta),\ \eta \in G\}$ of one-step probability matrices on S with $P(\eta) \equiv (p_{xy}(\eta))$ [13]. Here, under mild conditions on the mappings $\eta \rightarrow p_{xy}(\eta)$, the methodology of Metivier and Priouret is shown to lead to a result on a.s. convergence. Moreover the specific structure of the model at hand allows for great simplifications in their original arguments.

The paper is organized as follows: The set-up of the Stochastic Approximation algorithm studied here is described in Section 2, together with the basic results of the paper (Theorems 2.1–2.2). Section 3 presents several examples from the theory of Markov decision processes (MDPs) which illustrate the usefulness of the convergence theory established in this paper. These examples deal mainly with implementation issues which arise in the problem of "steering the cost to a given value" and in the theory of constrained MDPs; this provides the intuition behind the proposed adaptive algorithm. The required regularity properties are derived in Section 4 under minimal conditions and the main estimate that underlies the use of the ODE method is developed in Section 5 which contains the proof of Theorem 2.2.

A few words on the notation used throughout the paper: The set of all real numbers is denoted by $\mathbb{R}$, and $I(A)$ stands for the indicator function of a set A. Unless stated otherwise, the notation $\lim_n$ and $\underline{\lim}_n$ are understood with n going to infinity.

2. THE STOCHASTIC APPROXIMATIONS: SET-UP AND RESULTS

The Stochastic Approximation

Assume the state space to be a *finite* set S of cardinality d and let the parameter space G be a *compact convex* subset of $\mathbb{R}^p$. A family of stochastic matrices $\{P(\eta),\ \eta \in G\}$ on S is assumed given by specifying for all x and y in S, a Borel mapping $\eta \rightarrow p_{xy}(\eta)$ on G such that $0 \leq p_{xy}(\eta) \leq 1$ and $\sum_y p_{xy}(\eta) = 1$. Sometimes it will be convenient to use the notation $P(\eta)$ where $P(\eta) \equiv (p_{xy}(\eta))$.

All random variables (RVs) are defined on some sample space Ω which for convenience is taken to be the Cartesian product $\Omega := (S \times G)^\infty$ with generic element $\omega = (x_0, y_0, x_1, y_1, \ldots)$. The coordinate mappings $\{X_n, \eta_n,\ n = 0, 1, \ldots\}$ are defined by setting $X_n(\omega) := x_n$ and $\eta_n(\omega) := y_n$ for every ω in Ω and all $n = 0, 1, \ldots$. This sample space Ω is equipped with the σ-field $\mathcal{F} := \vee_{n=0}^{\infty} \mathcal{F}_n$ where $\mathcal{F}_n := \sigma\{X_0, \eta_0, X_1, \ldots, X_n, \eta_n\}$ for all $n = 0, 1, \ldots$, so that X_n and η_n are both RVs on $(\Omega, \mathcal{F})$.

The Stochastic Approximations of interest in this paper is the algorithm that produces the G-valued iterates $\{\eta_n,\ n = 0, 1, \ldots\}$ through the recursion

$$\eta_{n+1} = \Pi_G \left\{ \eta_n + a_n f(\eta_n, X_{n+1}) \right\} \qquad n = 0, 1, \ldots (2.1)$$
$$\eta_0 \in G$$

where f is a Borel measurable mapping $G \times S \rightarrow \mathbb{R}^p$, and Π_G denotes the the nearest-point projection on G [8]. The sequence of step sizes $\{a_n,\ n = 0, 1, \ldots\}$ satisfies the usual conditions

$$0 < a_n \downarrow 0, \qquad \sum_{n=0}^{\infty} a_n = \infty, \qquad \sum_{n=0}^{\infty} a_n^2 < \infty. \qquad (2.2)$$

The probabilistic evolution of the state process $\{X_n,\ n = 0, 1, \ldots\}$ is characterized by

$$P[X_{n+1} = y | \mathcal{F}_n] = p_{X_n y}(\eta_n) \qquad n = 0, 1, \ldots (2.3)$$

for all y in S.

Let μ be a probability measure on S. The requirements (2.1)–(2.3) defining the Stochastic Approximation algorithm induce a unique probability measure P on $\mathcal{F}$ such that X_0 is distributed according to μ under P. Moreover, for every η in G, it is convenient to introduce a probability measure P^η on $\mathcal{F}$ such that X_0 is again distributed according to μ under P^η and

$$P^\eta[X_{n+1} = y | \mathcal{F}_n] = p_{X_n y}(\eta) \qquad n = 0, 1, \ldots (2.4)$$

for all y in S, while the iterates $\{\eta_n,\ n = 0, 1, \ldots\}$ are still governed by (2.1). The existence and uniqueness of the measures P and $\{P^\eta, \eta \in G\}$ follow from the Kolmogorov Extension Theorem. The expectation operators under P and P^η are denoted by E and E^η, respectively.

Note from (2.4) that for each η in G, the RVs $\{X_n,\ n = 0, 1, \ldots\}$ form a Markov chain under P^η. As customary this Markov chain is identified with its matrix $P(\eta)$ of one-step probabilities $(p_{xy}(\eta))$.

The assumptions

The purpose of this paper is to provide mild conditions under which the iterates $\{\eta_n,\ n = 0, 1, \ldots\}$ generated through (2.1) converge a.s. under P to a (non-random) limit, and to characterize this limit. The assumptions of interest are stated below as conditions (C1)-(C4).

(C1) For each η in G, under P^η, the RVs $\{X_n,\ n = 0, 1, \ldots\}$ form an *aperiodic* Markov chain with a *single* recurrent class;

(C2) For all x and y in S, the transition probabilities $\eta \rightarrow p_{xy}(\eta)$ are *Lipschitz continuous* on G

(C3) For all x in S, the mapping $G \rightarrow \mathbb{R}^p\ :\ \eta \rightarrow f(\eta, x)$ is *Lipschitz continuous* on G.

Note that the properties of Lipschitz continuity stated in (C2) and (C3) are independent of the norms equipping $\mathbb{R}$ and $\mathbb{R}^p$. For the sake of definiteness the discussion is carried out with the understanding that the Euclidean spaces considered here are all equipped with the Euclidean norm. To fix the notation, let $|x|$ denote the Euclidean norm of any element x any space $\mathbb{R}^k$.

Under (C1), for each η in G, the Markov chain $\{X_n,\ n = 0, 1, \ldots\}$ is positive recurrent under P^η (since S is finite), and therefore possesses a unique invariant measure $\pi(\eta) \equiv (\pi(\eta, x))$. Set

$$F(\eta) := \sum\nolimits_{x \in S} \pi(\eta, x) f(\eta, x) \tag{2.5}$$

with the obvious interpretation that $F(\eta)$ is the expectation of $f(\eta, X)$ where X denotes a generic S-valued RV distributed according to $\pi(\eta)$. Now for η and F in $\mathbb{R}^p$, define the projection on G of the vector (field) F originating at η by

$$\Pi_G(\eta, F) := \lim\nolimits_{h \downarrow 0} \frac{\Pi_G\{\eta + hF\} - \eta}{h}. \tag{2.6}$$

The ODE

$$\frac{d}{dt}\eta(t) = \Pi_G(\eta(t), F(\eta(t))), \quad t \geq 0, \quad \eta(0) \text{ in } G \tag{2.7}$$

is the one associated with the algorithm (2.1)-(2.2) by the Kushner–Clark lemma [8, Thm. 5.3.1, pp. 191]. The existence and uniqueness of solutions to (2.7) is readily guaranteed under the conditions (C1)-(C3). Indeed, by (4.10) and Theorem 4.3 the mapping $\eta \rightarrow F(\eta)$ is Lipschitz continuous on G in view of (C3), and it is a simple exercise to check that the mapping $\eta \rightarrow \Pi_G(\eta, F(\eta))$ is also Lipschitz continuous on G. It will be crucial to require some form of stability for the ODE (2.7). This is the content of assumption (C4).

(C4) The ODE (2.7) is (Liapunov) *asymptotically stable* in G, and its stable point is η^*.

Recall that an ODE is said to be Liapunov asymptotically stable in a region G, with η^* as the attracting point, if (i) starting at any point in G the solution converges to η^*, and (ii) for any

$\epsilon > 0$ there exists $\delta > 0$ such that starting in a δ-neighborhood of η^*, the solution remains in the ϵ-neighborhood of η^* for *all* $t \geq 0$.

The results

The main result of this paper can now be stated.

Theorem 2.1 *Under the assumptions (C1)-(C4), the sequence of iterates $\{\eta_n, n = 0, 1, \ldots\}$ converges a.s. under P, i.e.,*

$$\lim_n \eta_n = \eta^* \qquad P - a.s. \tag{2.8}$$

The approach adopted here for establishing the convergence (2.8) uses an ODE argument based on the deterministic lemma of Kushner and Clark [8] as presented by Metivier and Priouret in [13]. The key result for the analysis is probabilistic in nature and is given in the next proposition whose proof is delayed till Section 5. To state the result, consider the RVs $\{Y_n,\ n = 0, 1, \ldots\}$ given by

$$Y_n := f(\eta_n, X_{n+1}) - F(\eta_n) \qquad n = 0, 1, \ldots (2.9)$$

and for every $T > 0$, pose

$$m(n,T) := \max\{k > n : \sum_{i=n}^{k-1} a_i \leq T\} \ . \qquad n = 0, 1, \ldots (2.10)$$

Theorem 2.2 *Under the assumptions (C1)-(C3), the convergence*

$$\lim_n \left(\sup_{n \leq k \leq m(n,T)} | \sum_{i=n}^{k} a_i Y_i | \right) = 0 \qquad P - a.s. \tag{2.11}$$

takes place.

Proof of Theorem 2.1. As explained by Metivier and Priouret [13], the convergence (2.11) underlines the P-a.s. convergence of $\{\eta_n,\ n = 0, 1, \ldots\}$ to η^*. The reader is invited to consult [8,13] for a complete exposition of the arguments which are now briefly summarized: Interpolate the estimate sequence $\{\eta_n,\ n = 0, 1, \ldots\}$, say by a piecewise linear function $\eta^{(0)} : [0, \infty) \to \mathbb{R}^p$ anchored in η_n at time $t_n = \sum_{i=0}^{n-1} a_i$, i.e., $\eta^{(0)}(t_n) = \eta_n$ for all $n = 0, 1, \ldots$, and define a sequence of left shifts $\eta^{(n)}(t) = \eta^{(0)}(t - t_n)$ which bring the "asymptotic part" of $\{\eta_n,\ n = 0, 1, \ldots\}$ back to a neighborhood of the time origin.

Now observe from (2.9) that the recursion (2.1) can be written in the form

$$\eta_{n+1} = \Pi_G \left\{ \eta_n + a_n \left[Y_n + F(\eta_n) \right] \right\}. \qquad n = 0, 1, \ldots (2.12)$$

From any convergent subsequence $\{\eta^{(m)}(\cdot),\ m = 0, 1, \ldots\}$ a further convergent subsequence $\{\eta^{(m_p)}(\cdot),\ p = 0, 1, \ldots\}$ can then be extracted by standard boundedness and equicontinuity arguments. It is then easy to see from (2.11) and (2.12) that its limit $\eta(\cdot)$, and for that matter the limit of *any* convergent subsequent, satisfies the ODE (2.7) which is *asymptotically stable* with a *unique* stable point η^*, as a consequence of (C4).

A simple shifting argument now implies $\eta(t) = \eta^*$ for all $t \geq 0$ and this completes the proof. These arguments are now standard and are omitted here in the interest of brevity. □

3. APPLICATIONS AND EXAMPLES

Many questions concerning MDPs can be reduced to searching for Markov stationary policies which satisfy certain constraints (or optimality) conditions. However, the resulting Markov stationary policies are usually *not* readily *implementable* [10], sometimes in spite of strong structural properties. This is so because the values of the model parameters may not be available [9,17], and even if they were available, the policy may still not be implementable due to computational difficulties inherent to its definition [17]. In some cases, these difficulties can be alleviated by considering alternatives based on a Stochastic Approximation algorithm of the type (2.2). This point is now developed in this section.

The MDP model

To set up the discussion, consider an MDP (S, U, P) as defined in the literature [16] where the state space S is a finite set and the action space U is an arbitrary measurable space. The one-step transition mechanism P is defined through the one-step transition probability functions $p_{xy}(\cdot) : U \to \mathbb{R}$ which are assumed to be *Borel* measurable and to satisfy the standard properties $0 \leq p_{xy}(u) \leq 1$ and $\sum_y p_{xy}(u) = 1$ for all x and y in S, and all u in U. The space of probability measures on U (when equipped with its natural Borel σ-field) is denoted by $\mathbb{M}$.

Here the canonical sample space for the MDP (S, U, P) is the Cartesian product $\Omega := S \times (U \times S)^\infty$ with generic element $\omega = (x_0, u_0, x_1, \ldots)$. Set $U_n(\omega) := u_n$ and $X_n(\omega) := x_n$ for every ω in Ω and all $n = 0, 1, \ldots$. The sample space Ω is equipped with the σ-field $\mathcal{F} := \vee_{n=0}^{\infty} \mathcal{F}_n$ where $\mathcal{F}_n := \sigma\{H_n\}$ and $H_n := (X_0, U_0, X_1, \ldots, U_{n-1}, X_n)$ for all $n = 0, 1, \ldots$, so that U_n and X_n are both RVs.

An *admissible* control policy γ is defined as any collection $\{\gamma_n, n = 0, 1, \ldots\}$ of measurable mappings $\gamma_n : S \times (U \times X)^n \to \mathbb{M}$ with the interpretation that for all $n = 0, 1, \ldots$, $\gamma_n(\cdot; H_n)$ is the probability distribution of selecting the control value U_n given the feedback information H_n.

Denote the collection of all such admissible policies by Γ.

Let μ be a fixed probability distribution on S. For every admissible policy γ in Γ, the Kolmogorov Extension Theorem guarantees the existence and uniqueness of a probability measure P^γ on the σ-field $\mathcal{F}$ so that under P^γ, the RV X_0 has distribution μ and

$$P^\gamma[X_{n+1} = y \mid \mathcal{F}_n] = \int_U \gamma_n(du; H_n) p_{X_n y}(u) \qquad n = 0, 1, \ldots (3.1)$$

for all y in S. The expectation operator associated with γ is denoted by E^γ.

A policy γ in Γ is said to be a *Markov* or *memoryless* policy if there exists a family $\{g_n,\ n = 0, 1, \ldots\}$ of mappings $g_n : S \to \mathbb{M}$ such that $\gamma_n(\cdot; H_n) = g_n(\cdot; X_n)\ P^\gamma - a.s.$ for all $n = 0, 1, \ldots$ In the event the mappings $\{g_n,\ n = 0, 1, \ldots\}$ are all identical to a given mapping $g : S \to \mathbb{M}$, the Markov policy is termed ***stationary*** and is identified with the mapping g itself. Under any Markov stationary g, the state process $\{X_n,\ n = 0, 1, \ldots\}$ evolves according to a Markov chain with one-step transition probability matrix $P(g) = (p_{xy}(g))$ given by

$$p_{xy}(g) := \int_U p_{xy}(u) g(du, x) \tag{3.2}$$

for all x and y in S.

Steering the cost

For any mapping $c : S \to \mathbb{R}$, define the corresponding long-run average cost functional $J_c : \Gamma \to \mathbb{R}$ by posing

$$J_c(\gamma) := \underline{\lim}_n\ E^\gamma \left[\frac{1}{n+1} \sum_{i=0}^{n} c(X_i) \right] \tag{3.3}$$

for every admissible policy γ in Γ. The problem of interest here is to find a Markov stationary policy g such that $J_c(g) = V$ for some constant V determined possibly through design considerations. The discussion assumes the existence of two *implementable* Markov stationary policies $\overline{g}$ and $\underline{g}$ such that

$$J_c(\overline{g}) < V < J_c(\underline{g}), \tag{3.4}$$

i.e., the Markov stationary policy $\overline{g}$ (resp. $\underline{g}$) undershoots (resp. overshoots) the requisite performance level V. As discussed below, this situation arises naturally in the solution of constrained MDPs via Lagrange arguments.

For every η in the unit interval [0,1], the policy f^η obtained by simply randomizing between the two policies $\overline{g}$ and $\underline{g}$ with *bias* η is the Markov stationary policy determined through the mapping $g^\eta : S \rightarrow \mathbb{M}$ where

$$g^\eta(\cdot;x) := \eta\,\underline{g}(\cdot;x) + (1-\eta)\,\overline{g}(\cdot;x) \tag{3.5}$$

for all x in S. Note that for $\eta = 1$ (resp. $\eta = 0$), the randomized policy g^η coincides with $\underline{g}$ (resp. $\overline{g}$). Owing to the condition (3.4), if the mapping $\eta \rightarrow J_c(g^\eta)$ is *continuous* on the interval [0,1], then the equation

$$J_c(g^\eta) = V, \quad \eta \text{ in } [0,1] \tag{3.6}$$

has at least one solution, say η^*, and $g = g^{\eta^*}$ thus steers (3.3) to the value V.

The implementation problem

Solving the (highly) nonlinear equation (3.6) for the bias value η^* is usually a non-trivial computational task, even in the simplest of situations [14]. This computational problem is further compounded by the parameter uncertainties that are inherent in the modeling of any system. Despite these difficulties, as illustrated by the examples below it is often possible to determine $\underline{g}$ and $\overline{g}$. In that case, a direct solution of (3.6) may be avoided by using an alternate policy $\alpha = \{\alpha_n,\ n = 0,1,\ldots\}$ of the form

$$\alpha_n(\cdot;H_n) := \eta_n\,\underline{g}(\cdot;X_n) + (1-\eta_n)\,\overline{g}(\cdot;X_n) \qquad n = 0,1,\ldots \tag{3.7}$$

for some sequence of [0,1]-valued RVs $\{\eta_n,\ n = 0,1,\ldots\}$ which act as "estimates" for the bias value η^*. This policy α constitutes an acceptable implementation of g provided $J_c(\alpha) = J_c(g)$.

In many applications, the mapping $\eta \rightarrow J_c(g^\eta)$ is monotone, say monotone increasing for sake of definiteness. The search for η^* can then be interpreted as finding the zero of the monotone function $\eta \rightarrow J_c(g^\eta) - V$ and this brings to mind ideas from the theory of Stochastic Approximations. The Robbins-Monro version of these algorithms suggests that the sequence of bias values $\{\eta_n,\ n = 0,1,\ldots\}$ be generated through the recursion

$$\eta_{n+1} = \Pi_{[0,1]}\left\{\eta_n + a_n\bigl(V - c(X_{n+1})\bigr)\right\} \qquad n = 0,1,\ldots \tag{3.8}$$

with η_0 given in [0,1], where the sequence of step sizes $\{a_n,\ n = 0,1,\ldots\}$ satisfies the conditions (2.2). This scheme (3.6)–(3.8) can be interpreted either as an estimation procedure, where the estimated parameter is defined through (3.6), or as an adaptive implementation scheme, where the controls are generated "on line" through (3.8).

Theorem 2.1 easily applies to the algorithm (3.7)-(3.8) under fairly mild conditions. A possible set of conditions are the assumptions (D1)-(D2), where

(D1) Under each one of the policies $\overline{g}$ and $\underline{g}$, the RVs $\{X_n \ n = 0, 1, \ldots\}$ form an *aperiodic* Markov chain with a *single* recurrent class;

(D2) The mapping $[0,1] \rightarrow \mathbb{R} : \eta \rightarrow J_c(g^\eta)$ is *strictly monotone increasing*,

The main properties of the implementation α are summarized in Theorem 3.1.

Theorem 3.1 *Assume (3.4). Under the assumptions (D1)-(D2), the following hold.*

(i): *The equation*

$$J_c(g^\eta) = V, \quad \eta \text{ in } [0,1], \tag{3.9}$$

has a unique solution η^* *in* $(0,1)$;

(ii): *The sequence of estimates* $\{\eta_n, \ n = 0, 1, \ldots\}$ *is strongly consistent under* P^α, *i.e.,*

$$\lim_n \eta_n = \eta^* \qquad P^\alpha - a.s. \tag{3.10}$$

(iii): *The policies* g *and* α *achieve the same cost, i.e.,*

$$J_c(\alpha) = J_c(g) = V. \tag{3.11}$$

Proof. Since for all x and y in S, $p_{xy}(\eta) := p_{xy}(g^\eta) = \eta p_{xy}(\underline{g}) + (1-\eta) p_{xy}(\overline{g})$ for every η in $[0,1]$, the mapping $\eta \rightarrow p_{xy}(\eta)$ is linear (and thus Lipschitz continuous) on $[0,1]$. By Lemma 4.2 and (4.7), the mapping $[0,1] \rightarrow \mathbb{R} : \eta \rightarrow J_c(g^\eta)$ is Lipschitz continuous on $[0,1]$, so that (3.6) admits at least one solution η^* in view of (3.4) and exactly one solution by virtue of (D2).

The assumption (D1) implies that under each one of the policies g^η, $0 \leq \eta \leq 1$, the state sequence $\{X_n \ n = 0, 1, \ldots\}$ form an aperiodic Markov chain with a single recurrent class. Indeed, this follows readily from the definitions of irreducibility and aperiodicity once it is observed that if for some $k = 0, 1, \ldots$ and some pair of states x and y in S, either $p_{xy}^{(k)}(\overline{g}) > 0$ or $p_{xy}^{(k)}(\underline{g}) > 0$, then $p_{xy}^{(k)}(g^\eta) > 0$ for all $0 < \eta < 1$. Consequently $F(\eta) = V - J_c(g^\eta)$, $0 \leq \eta \leq 1$, by standard results from the theory of Markov chains [5]. By the strict monotonicity assumption (D2), the projected ODE (2.6) can now be reduced in this scalar situation to

$$\dot{\eta}(t) = V - J_c(g^{\eta(t)}), \quad t \geq 0, \quad \eta(0) \text{ in } [0,1]. \tag{3.12}$$

That this ODE is asymptotically stable and that η^* is its unique stable point, follows from (D2) and (i). Part (ii) is now an immediate consequence of Theorem 2.1.

The result (3.11) on the cost is a simple consequence of the parameter convergence (3.10) and of a generalization [18] of a result by Mandl [12]. □

If the mapping $\eta \rightarrow J_c(g^\eta)$ were monotone decreasing, then the Stochastic Approximation algorithm (3.8) would be modified by replacing $V - c(X_{n+1})$ with $c(X_{n+1}) - V$, and the assumption (D2) would be changed accordingly.

Constrained optimization

Constrained MDPs provide a rich class of situations where the ideas given above have an immediate application. Let c and d be two cost functions $S \rightarrow \mathbb{R}$, and let $J_c(\gamma)$ and $J_d(\gamma)$ denote the corresponding long-run average costs (3.3) incurred under an arbitrary policy γ in Γ. With $\Gamma_V := \{\gamma \in \Gamma : J_c(\gamma) \geq V\}$ for some V in $\mathbb{R}$, consider the *constrained optimization* problem

$$\text{Maximize} \quad J_d(\cdot) \quad \text{over} \quad \Gamma_V.$$

In the event $c \leq 0$ and $d \geq 0$, the problem has the natural interpretation of maximizing the reward subject to a bound on the cost. Assume henceforth that Γ_V is non-empty and strictly contained in Γ, so that the problem is feasible but not trivial.

Beutler and Ross [4] have shown that under mild recurrence conditions, if U is *compact* and if the mappings $u \rightarrow p_{xy}(u)$ are continuous for all x and y in S, then there exist two *Markov deterministic* policies $\overline{g}$ and $\underline{g}$ so that (3.4) holds. Moreover, if g^η is given by (3.5), then $\eta \rightarrow J_c(g^\eta)$ is continuous, and if η^* solves (3.6), then $g = g^{\eta^*}$ is a solution to the constrained optimization problem.

Applying Theorem 3.1, it follows that if $\eta \rightarrow J_c(g^\eta)$ satisfies condition (D2), then the policy α defined through (3.7)–(3.8) satisfies $J_c(\alpha) = J_c(g) = V$. Similarly, $J_d(\alpha) = J_d(g)$ and α solves the constrained optimization problem.

In applications arising in queuing models, the control policies $\underline{g}$ and $\overline{g}$ are often simple to obtain, as the following examples illustrate.

Flow control

Consider a discrete-time queue $M/M/1$ queue with a finite buffer of size M (including the customer in service). Service completions and arrivals are modeled by two independent Bernoulli sequences. The controller implements a flow control mechanism by deciding whether or not to admit an arriving customer into the queue, with the understanding that a rejected customer is lost. Let X_n denote the number of customers in the system at time $n = 0, 1, \ldots$.

Under a variety of cost structures c and d, the optimal policy for the constrained problem has been shown to be of threshold type [9,7], i.e., if an arriving customer finds x customers in the system, it is accepted if $x < L$ for some threshold L and it is rejected if $L < x$. If $x = L$, then a coin with bias η^* is flipped, and the arriving customer is accepted or rejected according to the outcome. Define $\overline{g}$ (resp. $\underline{g}$) to be the policy that rejects customers when $x \geq L$ (resp. $x \geq L + 1$ respectively), and define g^η through (3.5).

When the cost function c that defines the constraint is strictly monotone, coupling arguments can be used to show strict monotonicity of the cost function $\eta \to J_c(g^\eta)$. Consequently, the optimal policy for the constrained problem is obtained by solving (3.9), and the scheme (3.7)-(3.8) solves the constrained optimization problem. Note that this implementation is insensitive to small modeling errors and as long as the correct threshold value remains L, the optimal bias is estimated on-line.

In fact, it is possible to generate an estimate of the optimal threshold L so that a fully adaptive scheme is obtained in the sense that no a priori knowledge of the model parameters is required: To this end, consider the recursion

$$\eta_{n+1} = \Pi_{[0,M]} \left\{ \eta_n + a_n \big(V - c(X_{n+1}) \big) \right\} \qquad n = 0, 1, \ldots (3.13)$$

with the interpretation that at time n, the threshold value $\lfloor \eta_n \rfloor$ (i.e., the largest integer value in η_n) and the bias $\eta_n - \lfloor \eta_n \rfloor$ are used. The monotonicity of $J_c(g^\eta)$ is again established through coupling arguments, whereas the Lipschitz continuity of the one-step transition probabilities are seen to hold by direct inspection. The result (3.11) is then established through an extension [18] of Mandl's result [12].

Resource allocation

As a final example, consider a discrete-time system of K infinite-capacity queues that compete for the service attention of a a single resource or server. The assumptions are the ones used in [2,3], i.e., service completions are modeled by Bernoulli RVs which are independent of the i.i.d. arrival batch process. Altman and Shwartz [1], and Nain and Ross [14] have studied the situation where the costs c and d are positive and linear in the queue sizes. It follows from their results that the optimal control in the presence of a constraint is obtained by randomizing (as in (3.5)) between two strict priority policies. In that case the policies $\underline{g}$ and $\overline{g}$ are easy to find in terms of the problem parameters. However, evaluating the optimal randomization bias η^* is computationally prohibitive since calculating $J_c(g^\eta)$ for $0 < \eta < 1$ involves solving a Riemann-Hilbert problem.

In this case, the state space S is not finite, so that Theorem 3.1 does not apply, but the scheme (3.7)-(3.8) is still of interest. For the case $K = 2$, Shwartz and Makowski [17] have obtained the results of Theorem 3.1 for this system, but where the convergence (3.10) holds in probability, rather than in the a.s. sense. However, the basic ideas of the present paper can be extended to this countable state system under appropriate moments conditions on the model data. This was done in [11] by Makowski and Shwartz who developed a method for proving a.s. convergence. The steering property (3.11) is established there via the results of [18].

4. SOME REGULARITY RESULTS

The proof of the convergence (2.8) is based on the ODE method as presented by Metivier and Priouret [13]. This approach hinges crucially on the fact that several quantities of interest are *Lipschitz* continuous (in the variable η) and it is the purpose of this section to establish the requisite regularity properties in some detail. In what follows, it will be convenient to view any mapping $f : S \to \mathbb{R}$ as a $d \times 1$ column vector $(f(x))$ (still denoted by f). Therefore, with this convention, any mapping $f : S \to \mathbb{R}^p$ can be represented as a $d \times p$ matrix $(f_1, \ldots, f_p)$. Also, let I_d denote the $d \times d$ identity matrix and let 0_d stand for the $1 \times d$ row vector with zero entries. Similarly, any mapping $f : G \times S \to \mathbb{R}$ can be viewed as a mapping $f : G \to \mathbb{R}^d$ through the convention $f(\eta) \equiv (f(\eta, x))$ introduced earlier. A similar convention is used to represent mappings $f : G \times S \to \mathbb{R}^p$.

Under (C1), the Markov chain $P(\eta)$ is *positive recurrent* for all η in G (since S is finite) and its *unique* invariant measure $\pi(\eta)$ is interpreted as a $1 \times d$ row vector $(\pi(\eta, x))$. It is well known that this invariant vector $\pi(\eta)$ is the *unique* solution to the system of equations

$$\pi = \pi P(\eta), \quad \pi e_d = 1 \tag{4.1}$$

in the variable $\pi = (\pi(x))$ in $\mathbb{R}^{1\times d}$ with e_d denoting the $d \times 1$ column vector with all entries equal to unity.

The next Lemma is useful for establishing the required regularity results. Throughout the discussion, the Lipschitz property of a matrix-valued mapping is understood entrywise.

Lemma 4.1 *If the mapping $G \to \mathbb{R}^{d\times d} : \eta \to A(\eta) = (A_{xy}(\eta))$ is Lipschitz with the property that the inverse $A^{-1}(\eta)$ of $A(\eta)$ exists for every η in G, then the mapping $G \to \mathbb{R}^{d\times d} : \eta \to A^{-1}(\eta)$ is Lipschitz on G.*

Proof. By standard results from Linear Algebra, there exist $d^2 + 1$ polynomial functions $r_0 : \mathbb{R}^{d^2} \rightarrow \mathbb{R}$ and $r_{xy} : \mathbb{R}^{d^2} \rightarrow \mathbb{R}$, with x and y ranging in S, in d^2 variables $A = (A_{xy})$ such that

$$A^{-1}(\eta)_{xy} = \frac{r_{xy}(A(\eta))}{r_0(A(\eta))} \tag{4.2}$$

for all x and y in S and all η in G. Here, these polynomial functions are of degree at most d and the relation $r_0(A(\eta)) = \det A(\eta) \neq 0$ holds for all η in G.

It now follows from the expression (4.2) that the mapping $\eta \rightarrow A^{-1}(\eta)_{xy}$ is *rational* for all x and y in S, thus locally Lipschitz at each point of G, except possibly at a finite number of points where the function may exhibit poles. However, $r_0(A(\eta))$ is Lipschitz in η and has no zero, so that the assumed Lipschitz continuity of the mapping $\eta \rightarrow A(\eta)$ precludes the existence of poles for each one of the mappings $\eta \rightarrow A^{-1}(\eta)_{xy}$ for all x and y in S. The result now follows from the fact that the function $[a, b] \rightarrow \mathbb{R} : x \rightarrow x^{-1}$ is Lipschitz continuous whenever $a > 0$ and that a product of Lipschitz continuous functions is also Lipschitz continuous. □

The smoothness of the components of $\pi(\eta)$ can now be investigated.

Lemma 4.2 *Under (C1)-(C2), the mapping $G \rightarrow \mathbb{R} : \eta \rightarrow \pi(\eta, x)$ is Lipschitz continuous for every x in S.*

Proof. The equations (4.1) satisfied by the invariant vector can be rewritten more compactly as

$$\pi Q(\eta) = [0_d \quad 1] \tag{4.3}$$

where $Q(\eta)$ is the $d \times (d+1)$ matrix given by

$$Q(\eta) := [I_d - P(\eta) \quad e_d]. \tag{4.4}$$

Consider the $d \times d$ matrix $\tilde{Q}(\eta)$ obtained from $Q(\eta)$ by removing its first column. Since the invariant measure is uniquely determined by (4.1), it is plain that $\pi(\eta)$ is the *unique* solution to the vector equation $\pi\tilde{Q}(\eta) = [0_{d-1} \quad 1]$ with an obvious interpretation for 0_{d-1}. Consequently $\tilde{Q}(\eta)$ is *invertible* and

$$\pi(\eta) = [0_{d-1} \quad 1]\tilde{Q}(\eta)^{-1} . \tag{4.5}$$

The mapping $\eta \rightarrow \tilde{Q}(\eta)$ is clearly Lipschitz on G due to (C2) and the result readily follows from Lemma 4.1. □

It is worth pointing out that under (C1), the relation

$$\lim_n E^\eta \left[\frac{1}{n+1} \sum_{i=0}^{n} 1[X_i = x] \right] = \pi(\eta, x) \tag{4.6}$$

holds for all x in S (independently of the initial distribution) by the standard Mean Ergodic Theorem for finite state Markov chains. Consequently, with $f : G \times S \to \mathbb{R}^p$ appearing in (2.1), the definitions (2.5) and (3.3) entail

$$J_{f_k}(\eta) = \lim_n E^\eta \left[\frac{1}{n+1} \sum_{i=0}^{n} f_k(\eta, X_i) \right] = \sum\nolimits_x \pi(\eta, x) f_k(\eta, x) =: F_k(\eta) \tag{4.7}$$

for all $1 \le k \le p$. The notation $F(\eta) = (F_1(\eta), \ldots, F_p(\eta))$ is used from now on.

Of interest here are the *Poisson* equations associated with the Markov chains $P(\eta)$, η in G, with forcing function $f : G \times S \to \mathbb{R}^p$. More precisely, for each η in G, a mapping $h : S \to \mathbb{R}^p$ and a vector J (in $\mathbb{R}^p$) solve the Poisson equation associated with $P(\eta)$ and forced by $f(\eta)$ if

$$h_k(x) + J_k = f_k(\eta, x) + \sum\nolimits_y p_{xy}(\eta) h_k(y), \quad 1 \le k \le p \tag{4.8a}$$

for all x in S, or in equivalent matrix form,

$$h_k + J_k e_d = f_k(\eta) + P(\eta) h_k, \quad 1 \le k \le p \tag{4.8b}$$

It is clear that if the pair (J, h) solves (4.8) so does $(J, h + e_d a)$ for every $1 \times p$ row vector a. Moreover, it is well known that if the pairs (J, h) and $(\tilde{J}, \tilde{h})$ both solve (4.8), then

$$J_k = \tilde{J}_k = \lim_n E^\eta \left[\frac{1}{n+1} \sum_{i=0}^{n} f_k(\eta, X_i) \right], \quad 1 \le k \le p \tag{4.9}$$

and $h - \tilde{h}$ is constant on the recurrent classes of $P(\eta)$.

It is plain that the solutions to the Poisson equation (4.8) depends on η. The remainder of this section is devoted to the study of the regularity properties of these solutions as a function of η. As pointed out earlier, the Markov chain $P(\eta)$ has a single positive recurrent class under (C1) (for each η in G), in which case the Poisson equation (4.8) has exactly one solution $(J(\eta), h(\eta))$ where $h(\eta) : S \to \mathbb{R}^p$ is determined up to an additive constant vector [19, Thm. 4.1]. A particular representative, still denoted $h(\eta)$, is now described. Before giving this definition, it is convenient to observe that

$$J_k(\eta) = \lim_n E^\eta \left[\frac{1}{n+1} \sum_{i=0}^{n} f_k(\eta, X_i) \right] = F_k(\eta), \quad 1 \leq k \leq p \tag{4.10}$$

as a result of (4.7) and (4.9).

For each η in G, define the stochastic matrix $P^*(\eta)$ by

$$P^*(\eta) := \lim_n \frac{1}{n+1} \sum_{i=0}^{n} P(\eta)^i. \tag{4.11}$$

This limit exists under (C1) by virtue of elementary results in the theory of Markov chains [5]. Since $P(\eta)$ has a single recurrent class, it is plain from (4.6) that all the rows of $P^*(\eta)$ are identical to $\pi(\eta)$, so that

$$P^*(\eta) = e_d \pi(\eta) \tag{4.12}$$

for all η in G.

It is now a simple exercise to see that the eigenvectors of $P^*(\eta)$ coincide with those of $P(\eta)$, and that the matrix $G(\eta) := P(\eta) - P^*(\eta)$ has spectral radius *strictly less* than unity, whence $I_d - G(\eta)$ is *invertible.* For all η in G, the mapping $h(\eta) : S \to \mathbb{R}^p$ is now defined by

$$h_k(\eta) := [I_d - G(\eta)]^{-1}[I_d - P^*(\eta)] f_k(\eta), \quad 1 \leq k \leq p. \tag{4.13}$$

Simple algebraic manipulations show that the pair $(J_k(\eta), h_k(\eta))$ given by (4.10) and (4.13) solves the Poisson equation (4.8), since $J_k(\eta) e_d = e_d \pi(\eta) f_k(\eta) = P^*(\eta) f_k(\eta)$ by virtue of (4.7) and (4.12).

Theorem 4.3 *Under the assumption (C1)-(C3), the solution pair to the Poisson equation (4.8) given by (4.10) and (4.13) is Lipschitz on G, i.e., the mappings $G \to \mathbb{R}^p : \eta \to J(\eta)$ and $G \to \mathbb{R}^p : \eta \to h(\eta, x)$, with x ranging over S, are all Lipschitz continuous.*

Proof. Since S is finite, the Lipschitz continuity of the mapping $\eta \to J(\eta)$ is an immediate consequence of Lemma 4.2 in view of (4.7) and (4.10).

The matrix-valued function $\eta \to P^*(\eta)$ is Lipschitz on G as a result of the representation (4.12) and of Lemma 4.2. It is now plain that the mappings $\eta \to I_d - P^*(\eta)$ and $\eta \to I_d - G(\eta)$ are both Lipschitz continuous on G, and the result now follows from Lemma 4.1 since the product of Lipschitz functions is clearly Lipschitz continuous. □

As a consequence of Theorem 4.3, since S is finite, there exists a positive constant K such that

$$\left|J(\eta) - J(\tilde{\eta})\right| \leq K|\eta - \tilde{\eta}| \qquad \text{and} \qquad \sup_x |h(\eta, x) - h(\tilde{\eta}, x)| \leq K|\eta - \tilde{\eta}| \tag{4.14}$$

for all η and $\tilde{\eta}$ in G.

5. A PROOF OF THEOREM 2.2

This section is devoted to the proof of the a.s. convergence result (2.11). It is plain from Theorem 4.3 that for each x in S, the mapping $\eta \to h(\eta, x)$ is continuous on the compact set G, thus bounded and therefore

$$B := \sup_\eta \sup_x \mid h(\eta, x) \mid < \infty \tag{5.1}$$

since S is finite. Moreover, the Poisson equation (4.8) easily implies that

$$E^\eta[h(\eta, X_{n+1}) \mid \mathcal{F}_n] = h(\eta, X_n) + J(\eta) - f(\eta, X_n) \qquad n = 0, 1, \ldots (5.2)$$

for all η in G, whence

$$\begin{aligned} &\mid E^\eta[h(\eta, X_{n+1}) \mid \mathcal{F}_n] - E^{\tilde{\eta}}[h(\tilde{\eta}, X_{n+1}) \mid \mathcal{F}_n] \mid \\ = &\mid h(\eta, X_n) - h(\tilde{\eta}, X_n) + J(\eta) - J(\tilde{\eta}) \mid \ \leq \ 2K \mid \eta - \tilde{\eta} \mid \end{aligned} \qquad n = 0, 1, \ldots (5.3)$$

for some $K > 0$ by making use of (4.14).

It follows from (2.9), (4.8) and (4.10) that

$$\begin{aligned} Y_n &= f(\eta_n, X_{n+1}) - J(\eta_n) \\ &= h(\eta_n, X_{n+1}) - E^{\eta_n}[h(\eta_n, X_{n+2}) \mid \mathcal{F}_{n+1}] \end{aligned} \qquad n = 0, 1, \ldots (5.4)$$

Set

$$Z_n^{(1)} := h(\eta_n, X_{n+1}) - E^{\eta_n}[h(\eta_n, X_{n+1}) \mid \mathcal{F}_n] \tag{5.5a}$$

$$Z_n^{(2)} := E^{\eta_n}[h(\eta_n, X_{n+1}) \mid \mathcal{F}_n] - E^{\eta_{n+1}}[h(\eta_{n+1}, X_{n+2}) \mid \mathcal{F}_{n+1}] \tag{5.5b}$$

and

$$Z_n^{(3)} := E^{\eta_{n+1}}[h(\eta_{n+1}, X_{n+2}) \mid \mathcal{F}_{n+1}] - E^{\eta_n}[h(\eta_n, X_{n+2}) \mid \mathcal{F}_{n+1}] \tag{5.5c}$$

for all $n = 0, 1, \ldots$. Define the RVs $\{S_n^{(k)},\ n = 0, 1, \ldots\}$ for all $k = 1, 2, 3$, by posing

$$S_n^{(k)} = \sum_{i=0}^{n-1} a_i Z_i^{(k)} \qquad n = 1, 2, \ldots (5.6)$$

with $S_0^{(1)} = S_0^{(2)} = S_0^{(3)} = 0$. Since $Y_n = Z_n^{(1)} + Z_n^{(2)} + Z_n^{(3)}$ for all $n = 0, 1, \ldots$, it suffices to show that

$$\lim_n \left(\sup_{n \leq \ell \leq m(n,T)} \mid \sum_{i=n}^{\ell} a_i Z_i^{(k)} \mid \right) = 0 \qquad P - a.s. \tag{5.7}$$

for all $T > 0$ and all $k = 1, 2, 3$.

It is plain that the RVs $\{Z_n^{(1)},\ n = 0, 1, \ldots\}$ form a $(P, \mathcal{F}_n)$ martingale-difference (taking values in $\mathbb{R}^p$), whence $\{S_n^{(1)},\ n = 0, 1, \ldots\}$ is a zero mean $(P, \mathcal{F}_n)$-martingale. Routine calculations show that

$$\sup_n E[\mid S_n^{(1)} \mid^2] = \sup_n E\left[\sum_{i=0}^{n-1} a_i^2 \mid Z_i^{(1)} \mid^2\right] \leq 4B^2 \sum_{i=0}^{\infty} a_i^2 \tag{5.8}$$

upon using (5.1) and (2.2), and the $(P, \mathcal{F}_n)$-martingale $\{S_n^{(1)},\ n = 0, 1, \ldots\}$ is thus uniformly integrable under P. By the Martingale Convergence Theorem, the RVs $\{S_n^{(1)},\ n = 0, 1, \ldots\}$ converge a.s. under P (to an a.s. finite limit), in which case they form a Cauchy sequence P-a.s. and (5.7) follows for $k = 1$.

To prove (5.7) for $k = 2$, note that for all $0 \leq n < \ell$, the relation

$$\begin{aligned} S_{\ell+1}^{(2)} - S_n^{(2)} &= \sum_{i=n}^{\ell} a_i Z_i^{(2)} \\ &= -\sum_{i=n}^{\ell} (a_{i-1} - a_i) E^{\eta_i}[h(\eta_i, X_{i+1}) \mid \mathcal{F}_i] \\ &\quad + a_{n-1} E^{\eta_n}[h(\eta_n, X_{n+1}) \mid \mathcal{F}_n] - a_\ell E^{\eta_{\ell+1}}[h(\eta_{\ell+1}, X_{\ell+2}) \mid \mathcal{F}_{\ell+1}] \end{aligned} \tag{5.9}$$

holds. It is now plain from (5.1), (2.2) and Jensen's inequality that

$$\mid S_{\ell+1}^{(2)} - S_n^{(2)} \mid \leq B \sum_{i=n}^{\ell} (a_{i-1} - a_i) + B(a_{n-1} + a_\ell) \tag{5.10}$$

$$\leq 2Ba_{n-1} \tag{5.11}$$

upon telescoping the terms in the first sum on the right handside of (5.10) and making use of the monotonicity of the sequence $\{a_n,\ n = 0, 1, \ldots\}$. The conclusion (5.7) for $k = 2$ is now immediate.

Finally for $k = 3$, note from (5.3) that

$$\mid Z_n^{(3)} \mid \leq 2K \mid \eta_n - \eta_{n+1} \mid . \qquad n = 0, 1, \ldots \tag{5.12}$$

Since the projection $x \to \Pi_G(x)$ is contracting on $\mathbb{R}^p$, the recursion (2.1) implies the inequality

$$| \eta_{n+1} - \eta_n | \leq a_{n+1} \mid f(\eta_n, X_{n+1}) \mid \leq \tilde{B} a_{n+1} \qquad n = 0, 1, \ldots (5.13)$$

with $\tilde{B} = \sup_\eta \sup_x \mid f(\eta, x) \mid$. Combining (5.12) and (5.13) leads to the inequality

$$\mid Z_n^{(3)} \mid \leq 2\tilde{B}K a_{n+1} \qquad n = 0, 1, \ldots (5.14)$$

and since the sequence $\{a_n, \ n = 0, 1, \ldots\}$ is decreasing, this leads to the bound

$$\sup_{n \leq \ell \leq m(n,T)} \left| \sum_{i=n}^{\ell} a_i Z_i^{(3)} \right| \leq \sum_{i=n}^{m(n,T)} a_i \mid Z_i^{(3)} \mid \leq 2\tilde{B}K \sum_{i=n}^{m(n,T)} a_i^2 \leq 2\tilde{B}K a_n (T + a_n) \, . \tag{5.15}$$

The convergence (5.7) now follows from (2.2). □

6. CONCLUDING REMARKS

The results of this paper can be given the interpretation either of an estimation procedure, where the estimated parameter is defined through $F(\eta) = 0$, or of an adaptive implementation scheme, as discussed in Section 3, where the controls are generated "on line" through (2.1). The paper concludes with several extensions of the results.

The results of this paper can be obtained under regularity conditions which are weaker than (C2)-(C3). One possible set of conditions under which the analysis carries through is stated below, where

(C2bis) The transition probabilities $\eta \to p_{xy}(\eta)$ are *Hölder continuous* for all x and y in S;

(C3bis) For all x in S, the mapping $\eta \to f(\eta, x)$ is *Hölder continuous* on G.

These conditions amount to requiring that there exist constants $K > 0$ and $0 < \beta \leq 1$ such that

$$|p_{xy}(f^\eta) - p_{xy}(f^{\tilde{\eta}})| \leq K|\eta - \tilde{\eta}|^\beta \tag{6.1}$$

for all x and y in S, with a similar condition for f.

In exact parallel with the developments of Sections 5 and 6, conditions (C1), (C2bis), (C3bis) and (C4) are sufficient to guarantee that

(i): For all x in S, the mapping $\eta \to \pi(\eta, x)$ is Hölder continuous with parameter β.

(ii): The mappings $\eta \to J(\eta)$ and $\eta \to h(\eta, x)$, with x ranging over S, are all Hölder continuous with parameter β.

(iii): If the iterates $\{\eta_n, n = 0, 1, \ldots\}$ are produced by (2.1), then (2.8) holds.

The proofs of (i)-(ii) are identical to the ones given for Lemma 4.2 and Theorem 4.3, respectively, upon observing that the class of Hölder continuous functions with parameter β is closed under addition and multiplication, and under composition with the function $x \to \frac{1}{x}$ on closed intervals which do not include 0. The proof of Theorem 2.2 carries over with a slight modification, namely that the last term in (5.3) and (5.12) needs to be changed to $2K|\eta - \tilde{\eta}|^\beta$. Modifying (5.14)-(5.15) accordingly, the last bound in (5.15) becomes $2\tilde{B}^\beta K a_n^\beta(T + a_n)$, which converges to zero due to (2.2).

If the regularity postulated in (C2bis)-(C3bis) is changed to continuous differentiability of order r (resp. analyticity), then the same remarks show that the smoothness in (i)-(ii) will then also be of order r (resp. analytic).

REFERENCES

[1] A. Altman and A. Shwartz, "Optimal priority assignment: a time-sharing approach," *IEEE Trans. Auto. Control*, to appear (1989).

[2] J.S. Baras, A.J. Dorsey and A.M. Makowski, "Two competing queues with geometric service requirements and linear costs: the μc-rule is often optimal," *Adv. Appl. Prob.* **17**, pp. 186-209 (1985).

[3] J.S. Baras, D.-J. Ma and A.M. Makowski, "K competing queues with geometric service requirements and linear costs: the μc-rule is always optimal," *Systems & Control Letters* **6**, pp. 173-180 (1985).

[4] F. Beutler and K. W. Ross, "Optimal policies for controlled Markov chains with a constraint," *J. Math. Anal. Appl.* **112**, pp. 236-252 (1985).

[5] K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, Second Edition, Springer-Verlag, New York (1967).

[6] E. G. Gladyshev, "On Stochastic Approximation," *Theo. Prob. Appl.* **10**, pp. 275-278 (1965).

[7] A. Hordijk and F. Spieksma, "Constrained admission control to a queuing system," *Adv. Appl. Prob.* **21** (1989).

[8] H. J. Kushner and D. S. Clark, *Stochastic Approximation for Constrained and Unconstrained Systems*, Applied Mathematical Sciences **26**, Springer-Verlag, Berlin (1978).

[9] D.-J. Ma and A. M. Makowski, "Optimality results for simple flow control problem," *Proceedings of the 26th IEEE Conference on Decision and Control*, Los-Angeles, California (1987).

[10] A. M. Makowski and A. Shwartz, "Implementation issues for Markov decision processes," pp. 323-337 in *Stochastic Differential Systems, Stochastic Control Theory and Applications*, Eds. W. Fleming and P.-L. Lions, The IMA Volumes in Mathematics and Its Applications **10**, Springer-Verlag, New York (1988).

[11] A. M. Makowski and A. Shwartz, "Analysis and adaptive control of a discrete-time single-server network with random routing," EE Pub. 719, Technion, Israel (1989).

[12] P. Mandl, "Estimation and control in Markov chains," *Adv. Appl. Prob.* **6**, pp. 40-60 (1974).

[13] M. Metivier and P. Priouret, "Applications of a Kushner and Clark lemma to general classes of stochastic algorithms," *IEEE Trans. Info. Theory* **AC-30**, pp. 140-150 (1984).

[14] P. Nain and K. W. Ross, "Optimal priority assignment with hard constraint," *IEEE Trans. Auto. Control* **AC-31**, pp. 883-888 (1986).

[15] H. Robbins and S. Monro, "A Stochastic Approximation method," *Ann. Math. Stat.* **22**, pp. 400-407 (1951).

[16] S. M. Ross, *Introduction to Stochastic Dynamic Programming*, Academic Press (1984).

[17] A. Shwartz and A. M. Makowski, "An optimal adaptive scheme for two competing queues with constraints," pp. 515-532 in *Proceedings of the 7th International Conference on Analysis and Optimization of Systems*, Eds. A. Bensoussan and J.-L. Lions, Springer Verlag Lecture Notes in Control and Information Sciences **83**, Antibes, France (1986).

[18] A. Shwartz and A. M. Makowski, "Comparing policies in Markov decision processes: Mandl's Lemma revisited," *Mathematics of Operations Research*, to appear (1989).

[19] A. Shwartz and A. M. Makowski, "On the Poisson equation for Markov chains," *Mathematics of Operations Research*, under revision (1987).