SRC TR 87-100

A Simple Problem of Flow Control II: Implementation of Threshold Policies Via Stochastic Approximations

by

Dye-Jyun Ma and Armand M. Makowski

DRAFT - May 1987

A SIMPLE PROBLEM OF FLOW CONTROL II: IMPLEMENTATION OF THRESHOLD POLICIES VIA STOCHASTIC APPROXIMATIONS

by

Dye-Jyun Ma[1] and Armand M. Makowski[2]

Electrical Engineering Department and Systems Research Center
University of Maryland, College Park, Maryland 20742

ABSTRACT

This paper considers a flow control model for discrete $M|M|1$ queues. The problem of implementing a given threshold policy via an adaptive policy is discussed in terms of an adaptive algorithm of the Stochastic Approximations type. Such an implementation problem for threshold policies typically arises when some of the model parameters are not exactly known. The proposed algorithm is extremely simple, easy to implement on-line, and requires no a priori knowledge of the actual values of the model parameters. Convergence of the algorithm and convergence of the long-run average cost under the adaptive policy are investigated. The obtained results apply easily to the optimal flow control studied in a companion paper [5].

[1] The work of this author was supported partially through NSF Grant ECS-83-51836 and partially through NSF Grant NSFD CDR-85-00108.

[2] The work of this author was supported partially through ONR Grant N00014-84-K-0614 and partially through a grant from AT&T Bell Laboratories. Please use the name of this author for any correspondence concerning this manuscript.

1. Introduction

Consider the following flow control model for discrete $M|M|1$ queues with infinite buffer capacity: Time is slotted so that the duration of a time slot coincides with the service requirement of any given customer. At the beginning of each time slot, the controller decides either to admit or to reject customers that arrive during that slot. This is done according to some prespecified mechanism on the basis of available (feedback) information. An admitted customer joins the queue while a rejected customer is immediately lost. During each time slot, a serviced customer (if any) may fail to complete service in that slot with a fixed positive probability, in which case it remains at the head of the line to await service in the next slot. The service failures are assumed independent from slot to slot, and independent of the arrival process. New customers arrive at the system one at a time according to a Bernoulli sequence.

This model was introduced in [5] and naturally arises in a variety of data communication systems. As congestion is experienced in such systems, it is natural to restrict access in order to guarantee certain levels of performance determined according to various design considerations [1]. In the companion paper [5], the selection of a flow control strategy was discussed with the objective of maximizing the throughput under the constraint that the long-run average number of customers in the system does not exceed a prespecified value, say V. By casting the problem as a constrained Markov decision process, a solution was shown to be of threshold type and to saturate the constraint. A threshold policy (L,η), with L in $\mathbb{N}$ and η in [0,1], has a simple structure in that at the beginning of each time slot, a new customer is accepted (resp. rejected) if the buffer content is strictly below L (resp. strictly above L), while if there are *exactly* L customers in the buffer, this new customer is accepted (resp. rejected) with probability η (resp. $1-\eta$).

In spite of its simple structure, a threshold policy (L,η) may not be implementable, as knowlege of L and/or η may not be readily available. To be more specific, as in [5], the threshold value L is often identified through the property that

$$J((L,0)) < V \leq J((L,1)) \tag{1.1}$$

where $J(\pi)$ denotes some long-run average cost functional associated with an admissible control policy π (see Section 2 for a precise definition) and V represents the constraint level. The desired threshold policy (L,η) is then obtained by appropriately choosing the bias value η

such that the constraint is met, i.e.,

$$J((L,\eta)) = V. \tag{1.2}$$

The cost functional $J((L,\eta))$ is a function of the model parameters, i.e, the arrival and service rates, and in the event of no exact knowledge of the model parameters, the exact values of L and/or η may not be known.

Motivated by this concern, *implementation* of a given threshold policy (L,η) is studied in this paper under the assumption that the threshold value L is readily available while the bias vlaue η has to be determined. This situation may arise, for instance, when only *approximate* values of the model parameters are available. To see this, if the cost functional $J(\pi)$ is rewritten as $J(\theta;\pi)$ to reflect this dependence on the model parameters θ, then it is possible for a wide range of values for θ to have

$$J(\theta;(L,0)) < V \leq J(\theta;(L,1)) \tag{1.3}$$

for the cost functional $J(\theta,\pi)$ is often *continuous* in θ. Those values θ satisfying (1.3), when used as if they were true model parameters, would naturally yield the same threshold value L.

In [7,9], various methods were proposed to overcome the implementation difficulties. In this paper, an alternative policy is generated through the Certainty Equivalence principle coupled with a specific estimation scheme for the *bias value* η. The proposed scheme is based on ideas from the theory of *Stochastic Approximations*. It is extremely simple, recursive and easy to implement on-line. It requires no a priori knowledge of the exact values of model parameters, as it avoids *direct* calculation of the control parameter η by solving the equation (1.2) even in the event of full knowledge of the model parameters. Convergence of the estimation scheme to the bias value η as well as convergence of the long-run average cost for a broad class of cost functionals under this adaptive policy are investigated. These results apply easily to the optimal flow control studied in [5]. The case where the model parameters are not known, is discussed in a companion paper [6], in which case both L and η are not available.

Although the implementation study of threshold polices given here applies only to this specific flow control model, the method of analysis and several ideas of the proof should prove useful in studying other related situations [8].

The paper is organized as follows: The model is described in Section 2. The Stochastic Approximations implementation policy is given in Section 3 where the main convergence results are summarized. Application of these results to the optimal flow control discussed in [5] is provided in Section 4, while various properties of threshold policies are outlined in Section 5 for future use. Section 6 is devoted to a general convergence result for the cost, and the convergence of the Stochastic Approximation algorithm is discussed in Sections 7-8. Finally, Section 9 gives some remarks on whether knowledge of the threshold value L can be relaxed and whether similar on-line estimation schemes can be used for the threshold value L.

A word on the notation: The set of real numbers is denoted by $\mathbb{R}$, with $\mathbb{R}_+$ denoting the set of non-negative real numbers. The set of all non-negative integers is denoted by $\mathbb{N}$, and for any x in $\mathbb{R}$, it is convenient to pose $\bar{x} = 1 - x$. The Kronecker delta $\delta(\bullet, \bullet)$ is defined as usual by $\delta(a, b) = 1$ if $a = b$ and $\delta(a, b) = 0$ otherwise. The characteristic function of any set E is denoed simply by $1[E]$.

2. Model

The model adopted here is the one given in the companion paper [5], and is briefly summarized for sake of convenience. The sample space Ω is taken to be the canonical space $\Omega := \mathbb{N} \times \left(\{0,1\}^3\right)^\infty$. The information spaces $\{\mathbb{H}_n\}_0^\infty$ are recursively generated by $\mathbb{H}_0 := \mathbb{N}$ and $\mathbb{H}_{n+1} := \mathbb{H}_n \times \{0,1\}^3$ for all n in $\mathbb{N}$, and with a slight abuse of notation, Ω is naturally identified with $\mathbb{H}_\infty$.

An element ω of Ω is viewed as a sequence $(x, \omega_0, \omega_1, \cdots)$ with x in $\mathbb{N}$ and ω_n in $\{0,1\}^3$ for all n in $\mathbb{N}$. Each block component ω_n is written in the form (u_n, a_n, b_n), with u_n, a_n and b_n being all elements in $\{0,1\}$. An element h_n in $\mathbb{H}_n$ is uniquely associated with the sample ω by $h_n := (x, \omega_0, \cdots, \omega_{n-1})$ with $h_0 := x$.

Let the sample $\omega = (x, \omega_0, \omega_1, \cdots)$ be realized. The initial queue size is set at x. During each time slot $[n, n+1)$, $a_n = 1$ (resp. a_n=0) indicates that a customer (resp. no customer) has arrived into the queue, $b_n = 1$ (resp. b_n=0) encodes a successful (resp. unsuccessful) completion of service in that slot, whereas control action u_n is selected at the beginning of the time slot $[n, n+1)$, with u_n=1 (resp. u_n=0) for admitting (resp. rejecting) the incoming customer during that slot. If x_n denotes the queue size at the beginning of the slot $[n, n+1)$, its successive values are determined through the recursion $x_{n+1} = x_n + u_n a_n - 1[x_n \neq 0] b_n$

with $x_0 := x$.

The coordinate mappings $\Xi, \{U(n)\}_0^\infty, \{A(n)\}_0^\infty$ and $\{B(n)\}_0^\infty$ are defined on the sample space Ω by posing $\Xi(\omega) := x$, $U(n,\omega) := u_n$, $A(n,\omega) := a_n$ and $B(n,\omega) := b_n$, while the information mappings $\{H(n)\}_0^\infty$ are given by $H(n,\omega) := (x, \omega_0, \omega_1, \cdots, \omega_{n-1}) := h_n$ for every ω in Ω and n in $\mathbb{N}$.

For each n in $\mathbb{N}$, let $\mathbb{F}_n$ be the σ-field generated by the mapping $H(n)$ on the sample space Ω. Clearly, $\mathbb{F}_n \subset \mathbb{F}_{n+1}$, and with standard notation, $\mathbb{F} := \vee_{n=0}^\infty \mathbb{F}_n$ is simply the natural σ-field on the infinite cartesian product $\mathbb{H}_\infty$ generated by the mappings Ξ and $\{U(n), A(n), B(n)\}_0^\infty$. Thus, on the space $(\Omega, \mathbb{F})$, the mappings $\Xi, \{U(n)\}_0^\infty$, $\{A(n)\}_0^\infty$, $\{B(n)\}_0^\infty$ and $\{H(n)\}_0^\infty$ are all random variables (RV) taking values in $\mathbb{N}$, $\{0,1\}$, $\{0,1\}$, $\{0,1\}$ and $\mathbb{H}_n$, respectively. The queue sizes $\{X(n)\}_0^\infty$ are $\mathbb{N}$-valued RV's which are defined recursively by

$$X(n+1) = X(n) + U(n)A(n) - 1[X(n) \neq 0]B(n) \qquad n = 0, 1, \cdots \quad (2.1)$$

with $X(0) := \Xi$. Each RV $X(n)$ is clearly $\mathbb{F}_n$-measurable.

An admissible policy π is defined as any collection $\{\pi_n\}_0^\infty$ of mappings $\pi_n : \mathbb{H}_n \to [0,1]$, with the interpretation that the potential arrival during the slot $[n, n+1)$ is admitted (resp. rejected) with probability $\pi_n(h_n)$ (resp. $1 - \pi_n(h_n)$) whenever the information h_n is available to the decision-maker. In the sequel, denote the collection of all such admissible policies by $\mathcal{P}$.

Let $q(\bullet)$ be a probability distribution on $\mathbb{N}$, and let λ and μ be fixed constants in $(0,1)$. Given any policy π in $\mathcal{P}$, there exists an unique probability measure P^π on $\mathbb{F}$, with corresponding expectation operator E^π, satisfying the requirements (R1)-(R3), where

(R1): For all x in $\mathbb{N}$,

$$P^\pi[\Xi = x] := q(x),$$

(R2): For all a and b in $\{0,1\}$,

$$\begin{aligned} P^\pi[A(n) = a, B(n) = b | \mathbb{F}_n \vee \sigma\{U(n)\}] &:= P^\pi[A(n) = a] P^\pi[B(n) = b] \\ &:= (a\lambda + \bar{a}\bar{\lambda})(b\mu + \bar{b}\bar{\mu}) \end{aligned} \qquad n = 0, 1, \cdots$$

(R3):

$$P^\pi[U(n) = 1 | \mathbb{F}_n] := P^\pi[U(n) = 1 | H(n)] := \pi_n(H(n)). \qquad n = 0, 1, \cdots$$

This notation is specialized to P_x^π and E_x^π, respectively, when $q(\bullet)$ is the point mass distribution at x in $\mathbb{N}$; it is plain that $P^\pi[A|X(0)=x] = P_x^\pi[A]$ for every A in $\mathbb{F}$.

It readily follows from (R1)-(R3) that under each probability measure P^π,

(P1): The $\mathbb{N}$-valued RV Ξ is independent of the sequences of RV's $\{A(n)\}_0^\infty$ and $\{B(n)\}_0^\infty$;

(P2): The sequences $\{A(n)\}_0^\infty$ and $\{B(n)\}_0^\infty$ of $\{0,1\}$-valued RV's are mutually independent Bernoulli sequences with parameters λ and μ, respectively;

(P3): The transition probabilities take the form

$$P^\pi[X(n+1)=y|\mathbb{F}_n] = p[X(n), y; \pi_n(H(n))] \qquad n=0,1,\cdots \quad (2.2)$$

where

$$p[x,y;\eta]:= \eta Q^1(x,y) + \bar{\eta} Q^0(x,y) \quad (2.3)$$

with

$$Q^i(x,y):= P^\pi[x+iA(n)-1(x\neq 0)B(n)=y], \quad i=0,1 \quad (2.4)$$

for all x and y in $\mathbb{N}$, and all η in [0,1].

The right-hand sides of (2.4) depend neither on n nor on the policy π owing to the assumptions (R1)-(R3) made earlier. Throughout this paper, the finite moment condition

$$E^\pi[\Xi] = \sum_{x=0}^{\infty} xq(x) < \infty \quad (2.5)$$

is assumed to hold for every π in $\mathcal{P}$.

A policy π in $\mathcal{P}$ is said to be a *Markov* policy if there exists a family $\{g_n\}_0^\infty$ of mappings $g_n: \mathbb{N} \to [0,1]$ such that $\pi_n(H(n)) = g_n(X(n))$ P^π-almost surely for all n in $\mathbb{N}$. In the event $g_n=g$ for all n in $\mathbb{N}$, the Markov policy π is called *stationary* and can be identified with the mapping g itself.

A policy π in $\mathcal{P}$ is said to be a *pure* (or *non-randomized*) policy if there exists a family $\{f_n\}_0^\infty$ of mappings $f_n: \mathbb{H}_n \to \{0,1\}$ such that $\pi_n(H(n)) = \delta(1, f_n(H(n)))$ P^π-a.s. for all n in $\mathbb{N}$. A *pure Markov stationary* policy π is thus fully characterized by a single mapping $f: \mathbb{N} \to \{0,1\}$.

A stationary policy g is said to be of *threshold* type if there exists a pair (L, η), with L an integer in $\mathbb{N}$ and η in $[0,1]$, such that

$$g(x) = \begin{cases} 1 & \text{if } x < L; \\ \eta & \text{if } x = L; \\ 0 & \text{if } x > L. \end{cases} \tag{2.6}$$

Such a *threshold* policy is denoted by (L, η), and remark that $(L,1) \equiv (L+1, 0)$. For convenience, the Markov stationary policy that admits every single customer, i.e., $g(x) = 1$ for all x in $\mathbb{N}$, is simply denoted by $(\infty, 1)$.

For any mapping $c: \mathbb{N} \to \mathbb{R}$, it is notationally convenient to pose

$$J^c(\pi) := \liminf_{n \uparrow \infty} \frac{1}{n+1} E^\pi \sum_{n=0}^{n} c(X(t)) \tag{2.7}$$

for every admissible policy π in $\mathcal{P}$ (whenever meaningful). Of interest in this paper are the mappings for which (2.7) is well-defined.

3. Problem statement and Stochastic Approximations

Let g denote a given threshold policy (L, η) with L in $\mathbb{N}$ and η in [0,1] held fixed throughout the discussion and assume the threshold value L to be available. The implementation problem of interest in this paper is to find a policy α in $\mathcal{P}$ which incurs the same cost as the policy g, i.e., $J^c(\alpha) = J^c(g)$, and which does not require *explicit* knowledge of the randomizing factor η. Such a policy α is called an *implementation* of g in the terminology of [9].

As discussed in [9], there are many possible implementations of g. The implementation scheme proposed in this paper is motivated by ideas of the theory of *Stochastic Approximations* and is of applicability in constrained flow control problems [Section 4]. With the notation $f^q := (L, q)$, $0 \le q \le 1$, observe that the threshold policy g can be interpreted as a simple randomization with bias η between the pure policies $f^0 := (L,0)$ and $f^1 := (L,1)$, i.e.,

$$g(x) = f^\eta(x) = \eta f^1(x) + \bar{\eta} f^0(x) \tag{3.1}$$

for all x in $\mathbb{N}$. Let V be a given constant and let r be a fixed mapping $\mathbb{N} \to \mathbb{R}$ such that the following assumptions (A1)-(A2) hold, where

(A1): *The mapping $r: \mathbb{N} \to \mathbb{R}$ is monotone, say increasing for sake of definiteness,*

(A2): *The mapping* $q \to J^r(f^q)$ *is continuous and strictly monotone increasing on the interval* [0,1], *with* $V = J^r(f^\eta) = J^r(g)$.

The cost functional J^r associated with the mapping r can be interpreted as the cost to be constrained according to some optimality criterion, or the quantity $V = J^r(g)$ can be viewed as one of particular interest toward which it is desirable to steer the system performance [7].

In the notation (3.1), the implementation policy α has the form

$$\alpha_n(H(n)) = \eta(n)f^1(X(n)) + \overline{\eta(n)}f^0(X(n)), \qquad n = 0, 1, \cdots (3.2)$$

where the bias estimates $\{\eta(n)\}_0^\infty$ are produced via the following Stochastic Approximation algorithm

$$\eta(n+1) = \left[\eta(n) + a_n(V - r(X(n+1))\right]_0^1. \qquad n = 0, 1, \cdots (3.3)$$

Here, $\eta(0)$ is chosen arbitrary in [0,1], and the convention $[x]_0^1 := 0 \vee (x \wedge 1)$ is enforced for all x in $\mathbb{R}$, while the step sizes $\{a_n\}_0^\infty$ form an $\mathbb{R}_+$-valued sequence which satisfies the conditions

$$0 < a_n \downarrow 0, \quad \sum_{i=0}^{\infty} a_i = \infty \quad \text{and} \quad \sum_{i=0}^{\infty} a_i^2 < \infty. \tag{3.4}$$

To show that $J^c(\alpha) = J^c(g)$ for a large class of mappings c under broad conditions, it is necessary to investigate the convergence of the estimates $\{\eta(n)\}_0^\infty$ to the bias value η under the adaptive policy α. The discussion of this convergence, which is available in Sections 7-8, is now summarized.

Theorem 3.1 *Assume the mapping* $r: \mathbb{N} \to \mathbb{R}$ *to be non-negative and to satisfy the assumptions* (A1)-(A2). *Whenever*

$$E^\pi[\Xi + \Xi r(\Xi) + r^2(\Xi)] < \infty \tag{3.5}$$

for all policy π *in* $\mathcal{P}$, *the sequence of biases* $\{\eta(n)\}_0^\infty$ *converges* P^α*- almost surely to the bias value* η.

By making use of Theorem 3.1, it is now possible to show that $J^c(\alpha) = J^c(g)$ by direct application of a general convergence result contained in Theorem 6.1.

Theorem 3.2 *Assume the conditions of Theorem 3.1 to hold. For any mapping* $c: \mathbb{N} \to \mathbb{R}$, *whenever the* $\mathbb{R}$*-valued RV's* $\{c(X(n))\}_0^\infty$ *form a uniformly integrable sequence under both*

probability measures P^α and P^g, the convergence

$$J^c(\alpha) = \lim_{n\uparrow\infty} \frac{1}{n+1} \sum_{t=0}^{n} c(X(t)) = J^c(g) \tag{3.6}$$

takes place in $L^1(\Omega, \mathbb{F}, P^\alpha)$, and consequently,

$$J^c(\alpha) = \lim_{n\uparrow\infty} \frac{1}{n+1} E^\alpha \sum_{t=0}^{n} c(X(t)) = J^c(g). \tag{3.7}$$

In particular, $J^r(\alpha) = J^r(g) = V$.

4. Application to optimal flow control

The results obtained in Theorems 3.1 and 3.2 can be readily applied to the optimal flow control problem considered in [5], where a policy that maximizes the throughput was sought under the constraint that the long-run average queue size does not exceed a certain value $V > 0$. For any admissible policy π in $\mathcal{P}$, the throughput $T(\pi)$ and the long-run average queue size $N(\pi)$ are naturally defined to be

$$T(\pi) := \liminf_{n\uparrow\infty} \frac{1}{n+1} E^\pi \sum_{t=0}^{n} \mu 1[X(t) \neq 0] \tag{4.1}$$

and

$$N(\pi) := \limsup_{n\uparrow\infty} \frac{1}{n+1} E^\pi \sum_{t=0}^{n} X(t), \tag{4.2}$$

respectively. With $\mathcal{P}_V := \{\pi \in \mathcal{P}:\ N(\pi) \leq V\}$, the constrained optimal control problem (P_V) is formally defined as

$$(P_V): \quad \text{maximize } T(\pi) \text{ over } \mathcal{P}_V.$$

The problem (P_V) always admits an optimal stationary policy for every $V > 0$; the optimality results, which were discussed at length in [5], are now sumarized.

Theorem 4.1 *If $N((\infty, 1)) \leq V$, the policy $(\infty, 1)$ solves the problem (P_V). If $N((\infty, 1)) > V$, then there exists a threshold policy $g = (L, \eta)$ which solves the problem (P_V) with $N(g) = V$.*

Only the situation $N((\infty,1)) > V$ is of interest here, in which case the Stochastic Approximation implementation α of g is defined by (3.3)-(3.4) with $r(x) = x$ for all x in $\mathbb{N}$, i.e., the bias estimates $\{\eta(n)\}_0^\infty$ are generated by

$$\eta(n+1) = \left[\eta(n) + a_n(V - X(n+1))\right]_0^1 \qquad n = 0, 1 \cdots (4.3)$$

with $\eta(0)$ arbitrary in [0,1]. It is now a simple exercise via Lemmas 5.1 and 5.4 to check that the assumptions of Theorems 3.1 and 3.2 hold for the costs (4.1)-(4.2), under the square-integrability of the RV Ξ. The optimality of the adaptive policy α for the problem (P_V) now follows easily.

Theorem 4.2 *Under a second moment assumption on the initial queue size, the adaptive policy α solves the problem (P_V) with $T(\alpha) = T(g)$ and $N(\alpha) = N(g) = V$.*

5. Properties of threshold policies

Various properties of threshold policies will be needed in the forthcoming discussion; they were obtained in the companion paper [5], and are summarized in this section for easy reference.

For each threshold policy $f^q := (L, q)$, $0 \leq q \leq 1$, the sequence $\{X(n)\}_0^\infty$ is a time-homogeneous Markov chain with state space $\mathbb{N}$ under the probability measure P^{f^q}. For ease of notation, rewrite $P^{f^q}, E^{f^q}, P_x^{f^q}$ and $E_x^{f^q}$ as P^q, E^q, P_x^q and E_x^q, respectively. This chain has a single ergodic set, namely $\{0, 1, \cdots, L+1\}$, and admits under P^q a unique invariant measure, which is denoted by $\mathbb{P}^q$ with corresponding expectation operator $\mathbb{E}^q$.

If X denotes a generic $\mathbb{N}$-valued RV, then the quantity $\mathbb{E}^q c(X)$ is always *finite* for any mapping $c : \mathbb{N} \to \mathbb{R}$. The following characterization combines Lemmas 5.1 and 5.2 of [5].

Lemma 5.1 *For any mapping $c : \mathbb{N} \to \mathbb{R}$, the convergence*

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{t=0}^{n} c(X(t)) = \mathbb{E}^q c(X) \qquad P^q - a.s. (5.1)$$

takes place, independently of the initial distribution. The convergence (5.1) also holds in $L^1(\Omega, \mathbb{F}, P^q)$, provided the sequence $\{c(X(n))\}_0^\infty$ is uniformly integrable under P^q. This is the case when the mapping c is monotone and the RV $c(\Xi)$ is integrable.

The solution to the Poisson equation associated with the cost function c and the threshold policy f^q will be particularly useful in what follows. As in [5], for each $i = 0, 1$, define $A^i(x) = x + iA - 1[x \neq 0]B$ for all x in $\mathbb{N}$, where A and B are generic elements in $\{A(n)\}_0^\infty$ and $\{B(n)\}_0^\infty$, respectively. The following lemma is a simple rephrasing of Lemma 5.4 of [5].

Lemma 5.2 *For any mapping $c: \mathbb{N} \to \mathbb{R}$ and any threshold policy $f^q, 0 \leq q \leq 1$, there always exist a scalar $J(q)$ and a mapping $h^q: \mathbb{N} \to \mathbb{R}$ such that*

$$h^q(x) + J(q) = c(x) + f^q(x)E[h^q(A^1(x))] + \overline{f^q(x)}E[h^q(A^0(x))] \tag{5.2}$$

for all x in $\mathbb{N}$. The quantity $J(q)$ is given by

$$J(q) = \lim_{n \uparrow \infty} \frac{1}{n+1} E_x^q[\sum_{t=0}^{n} c(X(t))] = \mathbb{E}^q c(X), \tag{5.3}$$

whereas the mapping $h^q: \mathbb{N} \to \mathbb{R}$ is unique up to an additive constant and is given by

$$h^q(x) = E_x^q[\sum_{t=0}^{\tau-1} c(X(t))] - E_x^q[\tau]J(q) \tag{5.4}$$

for all $x \neq L$ in $\mathbb{N}$ with $h^q(L) = 0$, where the $\mathbb{F}_n$-stopping time τ is defined as

$$\tau := \inf\{n \geq 0 : \ X(n) = L\}. \tag{5.5}$$

The invariant measure $\mathbb{P}^q$ was explicitly computed in Section 5 of [5], and exhibits the following property.

Lemma 5.3 *For $1 \leq k \leq L+1$, each one of the mappings $q \to \mathbb{P}^{f^q}[X \geq k]$ is continuously differentiable and strictly monotone increasing on the interval $[0, 1]$.*

The following characterization is obtained from Lemma 5.3 by arguments similar to the ones leading to Lemma 5.5 of [5].

Lemma 5.4 *If the mapping $c: \mathbb{N} \to \mathbb{R}$ is monotone increasing and if the RV $c(\Xi)$ is integrable, then the mapping $q \to J^c(f^q) = \mathbb{E}^q c(X)$ is continuously differentiable and monotone increasing on the interval $[0, 1]$. Furthermore, if the mapping c is not identically constant over $\{0, 1, \cdots, L+1\}$, then the mapping $q \to J^c(f^q) = \mathbb{E}^q c(X)$ is strictly monotone.*

6. A general convergence result for the costs

In this section, a general convergence result is obtained under the condition that an implementation α in $\mathcal{P}$ of the threshold policy g satisfies the following convergence condition (C) (with respect to g), where

(C): *The convergence*

$$\lim_{n\uparrow\infty} |\alpha_n(H(n)) - g(X(n))| = 0$$

takes place in probability under P^α.

To facilitate the discussion, it is convenient to introduce the conditions (H1)-(H3), where

(H1): *The sequence* $\{c(X(n))\}_0^\infty$ *of* $\mathbb{R}$*-valued RV's forms a uniformly integrable sequence under* P^g.

(H2): *The sequence* $\{c(X(n))\}_0^\infty$ *of* $\mathbb{R}$*-valued RV's forms a uniformly integrable sequence under* P^α.

(H3): *The sequence* $\{X(n)\}_0^\infty$ *of* $\mathbb{R}$*-valued RV's forms a uniformly integrable sequence under* P^α.

Note that under (H1)-(H2), $J^c(g)$ and $J^c(\alpha)$ are both well-defined and finite.

The main convergence result is now stated; although it is more general than is needed in this paper, the result has application in a companion paper [6] where various implementation policies are discussed.

Theorem 6.1 *Under the conditions* (H1)-(H3), *whenever the convergence condition* (C) *holds for the policy* α *with respect to the threshold policy* g, *the convergence*

$$J^c(\alpha) = \lim_{n\uparrow\infty} \frac{1}{n+1} \sum_{t=0}^{n} c(X(t)) = J^c(g) \tag{6.1}$$

takes place in $L^1(\Omega, \mathcal{F}, P^\alpha)$, *and consequently,*

$$J^c(\alpha) = \lim_{n\uparrow\infty} \frac{1}{n+1} E^\alpha \sum_{t=0}^{n} c(X(t)) = J^c(g). \tag{6.2}$$

A proof of Theorem 3.2 is now presented that makes use of Theorem 6.1.

A proof of Theorem 3.2

Theroem 3.1 readily implies that the convergence condition (C) holds for the policy α given by (3.3)-(3.4) with respect to g. On the other hand, Lemma 8.2 readily gives the condition (H3) of Theorem 6.1, i.e., the uniform integrability of the RV's $\{X(n)\}_0^\infty$ under P^α, and Theorem 6.1 thus applies to yield (3.6)-(3.7). That (3.6) also applies to the mapping r is an immediate consequence of Lemmas 8.2 and 5.1. □

The proof of Theorem 6.1 extends an argument due to Mandl [10] to the case of *unbounded* costs over countable state space and *randomized* policies. It is based on Theorem 6.3, where the convergence (6.1) is established for all *bounded* mappings. For the general case, the argument proceeds by considering the bounded mappings $c^B: \mathbb{N} \to \mathbb{R}$ defined by $c^B(x) = (c(x) \wedge B) \vee (-B)$, x in $\mathbb{N}$, for all $B > 0$, and Theorem 6.3 thus applies to each mapping c^B. The conditions (H1)-(H2) allow the convergence result for c^B to be carried over to the mapping c; its proof is similar to the one given by Shwartz and Makowski [12] for a competing queue problem, and is thus omitted here for sake of brevity. The interested reader is invited to consult the proof of Theorm 7.1 of [12, pp. 34] for a typical argument.

The discussion below is thus devoted to the case when the mapping c is bounded. Let the pair $(h, J) := (h^\eta, J(\eta))$ be the solution to the Poisson equation associated with the cost function c and the threshold policy $g = f^\eta$ as given in Lemma 5.2, the subscript η being omitted here for ease of notation.

As in [10], let the sequence $\{\Phi(n)\}_0^\infty$ and $\{Y(n)\}_0^\infty$ be defined by

$$\Phi(n) := E^\alpha[h(X(n+1))|\mathbb{F}_n] - E^g[h(X(n+1))|\mathbb{F}_n] \qquad n = 0, 1, \cdots (6.3)$$

and

$$Y(n+1) := h(X(n+1)) - E^\alpha[h(X(n+1))|\mathbb{F}_n] \qquad n = 0, 1, \cdots (6.4)$$

with $Y(0) := h(\Xi) - E^\alpha[h(\Xi)]$, respectively. By virtue of (2.2)-(2.4), (6.3) can be expressed as

$$\Phi(n) = \lambda[\alpha_n(H(n)) - g(X(n))][h(A^0(X(n)) + 1) - h(A^0(X(n)))] \qquad n = 0, 1, \cdots (6.5)$$

whereas (5.2) can be written in the form

$$\Phi(n) = c(X(n)) - J + h(X(n+1)) - h(X(n)) - Y(n+1). \qquad n = 0, 1, \cdots (6.6)$$

For $x > L$, the RV τ defined in (5.5) is distributed according to a negative Binomial under P_x^g, whence $E_x^g[\tau] = \frac{x-L}{\mu}$ [2, pp. 16], while by ergodicity $E_x^g[\tau] < \infty$ for $0 \leq x \leq L$. This fact,

when used on (5.4), shows that the mapping h exhibits linear growth, i.e., there exists some constant $K > 0$ such that

$$|h(x)| \leq K(1+x) \tag{6.7}$$

for all x in $\mathbb{N}$. This estimate (6.7) implies by (H3) and Lemma 5.1 that the RV's $\{h(X(n))\}_0^\infty$ are uniformly integrable under both P^α and P^g, whence the RV's $\{\Phi(n)\}_0^\infty$ and $\{Y(n)\}_0^\infty$ are well-defined. In the following lemma, useful bounds on $\{\Phi(n)\}_0^\infty$ and $\{Y(n)\}_0^\infty$ are further obtained by exploiting the Poisson equation (5.2) from which the relations

$$\mu(h(x+1) - h(x)) = c(x+1) - J \tag{6.8}$$

readily follow for all $x \geq L$.

Lemma 6.2 *There exist positive constants* K_1 and K_2 *such that*

$$|\Phi(n)| \leq K_1 |\alpha_n(H(n)) - g(X(n))| \Big(1 + |c(A^0(X(n)) + 1)|\Big) \tag{6.9}$$

and

$$|Y(n+1)| \leq K_2 \Big(1 + |c(A^0(X(n)) + 1)| + |c(X(n))| + |c(X(n)+1)|\Big) \tag{6.10}$$

for all n *in* $\mathbb{N}$.

Proof: The estimate

$$\begin{aligned} \mu|h(A^0(X(n)) + 1) - h(A^0(X(n)))| &= 1(A^0(X(n)) \geq L)|c(A^0(X(n)) + 1) - J| \\ &\quad + 1(A^0(X(n)) < L) 2 \max_{0 \leq x \leq L} |h(x)| \mu \end{aligned}$$

readily follows from (6.8), and (6.9) is now immediate from (6.5) for some constant $K_1 > 0$.

Since transitions can only take place into neighboring states, the relation (6.8) also implies that

$$\begin{aligned} \mu|h(X(n+1)) - h(X(n))| \leq 1(X(n)) > L)\{&1(X(n+1) = X(n) - 1)|c(X(n)) - J| \\ &+ 1(X(n+1) = X(n) + 1)|c(X(n) + 1) - J|\} \\ &+ 1(X(n) \leq L) 2 \max_{0 \leq x \leq L+1} |h(x)| \mu \end{aligned} \tag{6.11}$$

and (6.10) thus follows from (6.6) and (6.9) for some constant $K_2 > 0$. □

Since the mapping c is bounded, the sequences $\{\Phi(n)\}_0^\infty$ and $\{Y(n)\}_0^\infty$ are both *bounded* by (6.9)-(6.10), and thus well-defined. By using this fact coupled with condition (H3), it is now possible to show the convergence (6.1) for all the bounded costs.

Theorem 6.3 *Assume c to be bounded. Under the condition* (H3), *whenever the convergence condition* (C) *holds for the policy α with respect to g, the convergence*

$$J^c(\alpha) = \lim_{n\uparrow\infty} \frac{1}{n+1} \sum_{t=0}^{n} c(X(t)) = J^c(g) \tag{6.12}$$

takes place in $L^1(\Omega, \mathcal{F}, P^\alpha)$.

Proof: Iteration of (6.6) readily implies

$$\begin{aligned} & J + \frac{1}{n+1} \sum_{t=0}^{n} \Phi(t) + \frac{1}{n+1} \sum_{t=0}^{n} Y(t+1) \\ & = \frac{1}{n+1} \sum_{t=0}^{n} c(X(t)) + \frac{1}{n+1} h(X(n+1)) - \frac{1}{n+1} h(\Xi) \end{aligned} \tag{6.13}$$

for all n in $\mathbb{N}$. From (6.9), it is plain that

$$|\Phi(n)| < K_1(1+B)|\alpha_n(H(n)) - g(X(n))| \qquad n = 0, 1, \cdots \tag{6.14}$$

with $B := \sup_x |c(x)|$. By the convergence condition (C), the RV's $\{|\alpha_n(H(n)) - g(X(n))|\}_0^\infty$ converge to zero in probability under P^α, hence the RV's $\{\Phi(n)\}_0^\infty$ thus also converge to zero in probability under P^α owing to (6.14). The RV's $\{\Phi(n)\}_0^\infty$ being bounded, convergence in probability implies convergence in $L^1(\Omega, \mathcal{F}, P^\alpha)$ and $\lim_{n\uparrow\infty} E^\alpha|\Phi(n)| = 0$. Elementary results on Cesaro convergence immediately yield

$$\lim_{n\uparrow\infty} \frac{1}{n+1} E^\alpha \sum_{t=0}^{n} |\Phi(t)| = 0. \tag{6.15}$$

From (6.4), the RV's $\{Y(n)\}_0^\infty$ are seen to form a $(P^\alpha, \mathcal{F}_n)$-*martingale difference* sequence, and the estimate

$$E^\alpha[\sum_{n=1}^{\infty} \frac{Y^2(n)}{n^2}] \leq \sup_n E^\alpha[|Y(n)|^2] \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty. \tag{6.16}$$

holds owing to (6.10). By a martingale version of the Law of Large Numbers [10, Thm. 3], the convergence

$$\lim_{n\uparrow\infty} \frac{1}{n+1} \sum_{t=0}^{n} Y(t+1) = 0 \tag{6.17}$$

takes place P^α-almost surely and thus also in $L^1(\Omega, I\!F, P^\alpha)$ by the Bounded Convergence Theorem.

Uniform integrability of the RV's $\{h(X(n)\}_0^\infty$ under P^α impies that $\sup_n E^\alpha|h(X(n))| < \infty$, whence

$$\lim_{n\uparrow\infty} \frac{1}{n+1} E^\alpha|h(\Xi)| = \lim_{n\uparrow\infty} \frac{1}{n+1} E^\alpha|h(X(n+1))| = 0. \tag{6.18}$$

Since $J = J^c(g)$ as a combined result of (5.3) and Lemma 5.1 since the RV's $\{c(X(n))\}_0^\infty$ are bounded. The result now follows upon letting n go to infinity in (6.13) and collecting (6.15), (6.17)-(6.18). □

7. Discussion of the convergence for the estimates

The convergence of the estimates $\{\eta(n)\}_0^\infty$ to the bias value η is now investigated. A proof of Theorem 3.1 is now provided which uses a standard ODE argument based on the deterministic lemma of Kushner and Clark [3]. The presentation follows closely that of Metivier and Priouret [11].

For each $0 \leq q \leq 1$, pose

$$\delta(q) := V - J^r(f^q), \tag{7.1}$$

and define the RV's $\{\xi(n)\}_0^\infty$ and $\{d(n)\}_0^\infty$ by

$$\xi(n) := J^r(f^{\eta(n)}) - r(X(n+1)) \qquad n = 0, 1, \cdots \tag{7.2}$$

and

$$d(n) = \frac{1}{a_n}\Big[\eta(n+1) - [\eta(n) + a_n\delta(\eta(n)) + a_n\xi(n)]\Big]. \qquad n = 0, 1, \cdots \tag{7.3}$$

With this notation, the recursion (3.3) becomes

$$\eta(n+1) = \eta(n) + a_n\delta(\eta(n)) + a_n\xi(n) + a_n d(n). \qquad n = 0, 1, \cdots \tag{7.4}$$

It is plain that

$$|d(n)| \begin{cases} = 0, & \text{if } \ 0 \le \eta(n) + a_n[\delta(\eta(n)) + \xi(n)] \le 1; \\ \le |V - r(X(n+1))|, & \text{otherwise.} \end{cases} \tag{7.5}$$

By the very definition of the policy α defined by (3.3), the estimate

$$X(k) \le (L+1) \vee X(n) \qquad P^\alpha - a.s. \tag{7.6}$$

holds for all $k > n$, implying

$$X(n) \le (L+1) \vee \Xi \qquad P^\alpha - a.s. \tag{7.7}$$

for all n in $I\!N$, and therefore

$$\sup_n |d(n)| \le V + r(L+1) + r(\Xi) \qquad P^\alpha - a.s. \tag{7.8}$$

since the mapping r is non-negative and monotone increasing.

For every $T > 0$, pose

$$m(n,T) := \max\{k > n : \sum_{i=n}^{k-1} a_i \le T\}. \qquad n = 0, 1, \cdots \tag{7.9}$$

The next result is key in making the Lemma of Kushner and Clark useful for questions of almost sure convergence of Stochastic Approximation algorithms.

Theorem 7.1 *Under the assumptions of Theorem 3.1, the convergence*

$$\lim_{n \uparrow \infty} \Big(\sup_{n \le k \le m(n,T)} |\sum_{i=n}^{k} a_i \xi(i)| \Big) = 0 \qquad P^\alpha - a.s. \tag{7.10}$$

takes place for every $T > 0$.

The proof of Theorem 7.1 requires several technical lemmas and is delayed till the next section. The proof of Theorem 3.1 is now given below with the help of Theorem 7.1.

A proof of Theorem 3.1

The main idea consists in interpolating $\{\eta(n)\}_0^\infty$ and then defining a sequence of left shifts which bring the "asymptotic part" of $\{\eta(n)\}_0^\infty$ back to a neighborhood of the time origin. To

that end, define the increasing sequence $\{t_n\}_0^\infty$ by $t_n = \sum_{k=0}^{n-1} a_k$ for all $n \geq 1$ with $t_0 = 0$. For any sequence $\{x(n)\}_0^\infty$, the *linear* interpolation and the *right continuous step* interpolation of the functions taking value $x(n)$ at t_n are denoted by $l.i.[(t_n, x(n)), \bullet]$ and $s.i.[(t_n, x(n)), \bullet]$, respectively, and are defined on $[0, \infty)$ by

$$x^0(t) := l.i.[(t_n, x(n)), t] := \frac{(t - t_n)x(n+1) + (t_{n+1} - t)x(n)}{t_{n+1} - t_n}, \quad t_n \leq t \leq t_{n+1}$$

and

$$\bar{x}(t) := s.i.[(t_n, x(n)), t] := x(n), \quad t_n \leq t < t_{n+1}.$$

With this notation, the functions $\eta^0(\bullet)$, $M^0(\bullet)$, $D^0(\bullet)$ and $\bar{\eta}(\bullet)$ are defined on $[0, \infty)$ by

$$\eta^0(t) := l.i.[(t_n, \eta(n)), t], \quad M^0(t) := l.i.[(t_n, \sum_{i=0}^{n-1} a_i \xi(i)), t],$$

$$D^0(t) := l.i.[(t_n, \sum_{i=0}^{n-1} a_i d(i)), t] \quad \text{and} \quad \bar{\eta}(t) := s.i.[(t_n, \eta(n)), t].$$

For each $n \geq 1$, the functions $\eta^n(\bullet)$, $M^n(\bullet)$ and $D^n(\bullet)$ are the "left shifts" of $\eta^0(\bullet)$, $M^0(\bullet)$ and $D^0(\bullet)$, respectively, and are now defined on $\mathbb{R}$ by

$$\eta^n(t) := \begin{cases} \eta^0(t + t_n), & \text{if } t \geq -t_n; \\ \eta(0), & \text{if } t \leq -t_n, \end{cases}$$

$$M^n(t) := \begin{cases} M^0(t + t_n) - M^0(t_n), & \text{if } t \geq -t_n; \\ 0, & \text{if } t \leq -t_n, \end{cases}$$

and

$$D^n(t) := \begin{cases} D^0(t + t_n) - D^0(t_n), & \text{if } t \geq -t_n; \\ 0, & \text{if } t \leq -t_n. \end{cases}$$

From the relation (7.4), it follows that for all $t \geq 0$,

$$\eta^0(t) = \eta(0) + \int_0^t \delta(\bar{\eta}(s))ds + M^0(t) + D^0(t), \tag{7.11}$$

and the function $\eta^n(\bullet)$ thus satisfies the relation

$$\eta^n(t) = \eta^n(0) + \int_0^t \delta(\eta^n(s))ds + M^n(t) + D^n(t) + \epsilon^n(t), \tag{7.12}$$

for all $t \geq 0$, with

$$\epsilon^n(t) := \int_0^t \delta(\bar{\eta}(t_n + s))ds - \int_0^t \delta(\eta^n(s))ds. \tag{7.13}$$

Let $\mathcal{N}$ be a P^α-null set on which (7.6) or (7.10) fails, and fix an ω not in $\mathcal{N}$. The sequence $\{\eta^n(\bullet,\omega)\}_0^\infty$ is bounded and equicontinuous; the boundness is obvious from (3.3), while the equicontinuity follows from the fact that $\eta^0(\bullet,\omega)$ is (globally) Lipschitz owing to (7.8).

The relation (7.10) implies that for every $T > 0$,

$$\lim_{n\uparrow\infty} \sup_{t\leq T} |M^n(t,\omega)| = 0, \tag{7.14}$$

and $M^n(\bullet,\omega)$ thus converges to zero uniformly on finite intervals as $n \uparrow \infty$. Also, $\epsilon^n(\bullet,\omega)$ converges to zero as $n \uparrow \infty$ uniformly on finite intervals due to the continuity of the mapping δ, the boundness of $\{\eta(n)\}_0^\infty$ and the fact that the step sizes $a_n \downarrow 0$. Since $\{d(n,\omega)\}_0^\infty$ is a bounded sequence owing to (7.8), the reader will check that the function $D^0(\bullet,\omega)$ is (globally) Lipschitz, whence the sequence $\{D^n(\bullet,\omega)\}_0^\infty$ is equicontinuous.

By the Arzela-Ascoli Theorem, a convergent subsequence $\{(\eta^m(\bullet,\omega), D^m(\bullet,\omega))\}_1^\infty$ can be selected. Owing to (7.12), the corresponding limit point $(\eta(\bullet,\omega), D(\bullet,\omega))$ must satisfy the equation

$$\eta(t) = \eta(0) + \int_0^t \delta(\eta(s))ds + D(t) \tag{7.15}$$

with $\eta(t)$ in [0,1] for all $t \geq 0$.

If $\eta(t,\omega)$ lies in $(0,1)$ on some interval $[t_1,t_2]$, then $D(t,\omega) = 0$ on that interval because of the relation (7.5) and uniform convergence. On the other hand, if $\eta(t,\omega)$ lies in $\{0,1\}$ on some interval $[t_3,t_4]$, say $\eta(t,\omega) = 1$, then, for $0 \leq t_3 \leq t_3 + h \leq t_4 < \infty$, $h > 0$,

$$\eta(t_3+h,\omega) = \eta(t_3,\omega) + \int_{t_3}^{t_3+h} \delta(\eta(s,\omega))ds + D(t_3+h,\omega) - D(t_3,\omega), \tag{7.16}$$

which implies

$$D(t_3+h,\omega) - D(t_3,\omega) = -h\delta(1). \tag{7.17}$$

By assumption (A2), the mapping $q \to \delta(q) = V - J^r(f^q)$ is continuous and strictly monotone decreasing, thus with a single zero η in the interval [0,1]. If $\eta = 1$, then $\delta(1) = 0$ and the

right-hand side of (7.17) is equal to zero. If $\eta < 1$, then $\delta(q) < 0$ for $\eta < q \leq 1$, and this implies by virtue of (7.17) that $D(\bullet,\omega)$ will increase linearly on the interval $[t_3, t_4]$. But this is impossible, since for all m sufficiently large, $\eta^m(\bullet,\omega)$ is close to 1 on the interval $[t_3, t_4]$ and this implies that $D^m(t_3 + h, \omega) - D^m(t_3, \omega) \leq 0$ by the very definition of the RV's $\{d(n)\}_0^\infty$. The situation $\eta(t,\omega) = 0$ on the interval $[t_3, t_4]$ can be similarly discussed. It now follows from (7.15) that the limit $\eta(\bullet,\omega)$ satisfies the ODE

$$\dot{\eta}(t) = \delta(\eta(t)) = V - J^r(f^{\eta(t)}), \tag{7.18}$$

for all $t \geq 0$, with $\eta(0)$ in $[0,1]$. The ODE (7.18) is *asymptotically stable* and *any* one of its solutions $\eta(\bullet)$ converges *monotonically* to η, which is the *unique* solution of the equation $J^r(f^q) = V$, $0 \leq q \leq 1$.

A simple shifting argument can now be used to show that $\lim_{n\uparrow\infty} \eta(n,\omega) = \eta$. For every $T > 0$, define the process $\eta_T^n(\bullet,\omega)$ by $\eta_T^n(t,\omega) := \eta^n(t - T, \omega)$ for all t in $\mathbb{R}$. Assume that the subsequence $\{\eta^m(\bullet,\omega)\}_0^\infty$ converges, say to $\eta(\bullet,\omega)$, in which case

$$\lim_{m\uparrow\infty} \left(\eta^m(\bullet,\omega), \eta_T^m(\bullet,\omega)\right) = \left(\eta(\bullet,\omega), \eta_T(\bullet,\omega)\right) \tag{7.19}$$

also takes place. The limit $\eta_T(\bullet,\omega)$ satisfies the same ODE (7.18) with stable point η, and initial condition $\eta_T(0,\omega)$ in [0,1]. By stability, it is possible to choose T large enough so that $\eta_T(T,\omega)$ is arbitrarily close to η, i.e., $|\eta_T(T,\omega) - \eta| \leq \delta_T$ with $\lim_{T\uparrow\infty} \delta_T = 0$, and the choice of T can be made *independently* of the subsequence. From the obvious equality $\eta_T(T,\omega) = \eta(0,\omega)$, since the solution converges *monotonically*, it follows that $|\eta(t,\omega) - \eta| \leq \delta_T$ for all $t \geq 0$. Since the choice of T was arbitrary, therefore $\eta(t,\omega) \equiv \eta$, which implies that $\eta(n,\omega) \to \eta$ along the convergent subsequence. Any convergent subsequence of $\{\eta(n,\omega)\}_0^\infty$ thus converges to the same limit η, whence

$$\lim_{n\uparrow\infty} \eta(n,\omega) = \eta \tag{7.20}$$

for all ω not in the P^α-null set $\mathcal{N}$, and this establishes the theorem. □

8. A proof of Theorem 7.1

The proof uses ideas proposed by Metivier and Priouret [11] and is developed in a series of auxiliary lemmas. As in Section 6, the argument starts with a use of Lemma 5.2, with

$(h^q, J(q))$ denoting the solution to the Poisson equation associated with the cost r and the threshold policy f^q for each $0 \leq q \leq 1$.

Lemma 8.1 *There exists a positive constant K such that for all $0 \leq q, \tilde{q} \leq 1$, the relations*

$$\left|h^q(X(n+1)) - E^q[h^q(X(n+1))|\mathcal{F}_n]\right| \leq K(1+r(X(n))), \tag{8.1}$$

$$\left|E^q[h^q(X(n+1))|\mathcal{F}_n]\right| \leq K(1+X(n))(1+r(X(n))) \tag{8.2}$$

and

$$\left|E^q[h^q(X(n+1))|\mathcal{F}_n] - E^{\tilde{q}}[h^{\tilde{q}}(X(n+1))|\mathcal{F}_n]\right| \leq K(1+X(n))|q-\tilde{q}| \tag{8.3}$$

take place P^α-almost surely for all n in $\mathbb{N}$.

Proof: Key to the proof is the relation

$$E^q[h^q(X(n+1))|\mathcal{F}_n] = h^q(X(n)) + J(q) - r(X(n)) \tag{8.4}$$

which follows from (5.2). The reader will check that

$$\begin{aligned} &\left|h^q(X(n+1)) - E^q[h^q(X(n+1))|\mathcal{F}_n]\right| \\ &= \left|h^q(X(n+1)) - h^q(X(n)) - J(q) + r(X(n))\right|. \end{aligned} \tag{8.5}$$

Since $X(n+1) \leq X(n)$ P^α-a.s. on the set $[X(n) > L]$, an argument similar to the one leading to (6.11), yields the estimate

$$\left|h^q(X(n+1)) - h^q(X(n))\right| \leq M(1+r(X(n))) \tag{8.6}$$

for some constant $M > 0$, and (8.1) is now established by combining (8.5)-(8.6) and the fact $J(0) \leq J(q) \leq J(1)$.

From (7.7), it is plain that $X(n) \leq x \vee (L+1)$ under P_x^q for all n in $\mathbb{N}$ whereas (5.4) implies

$$|h^q(x)| \leq \Big(J(q) + r(x \vee (L+1))\Big) E_x^q[\tau].$$

The mapping r being non-negative and monotone increasing, the relation (8.2) follows readily from (8.4) since the mapping $x \to E_x^q[\tau]$ is at most linear.

Using (8.4) again, the reader will check that

$$\begin{aligned} &\left|E^q[h^q(X(n+1))|\mathbb{F}_n] - E^{\tilde{q}}[h^{\tilde{q}}(X(n+1))|\mathbb{F}_n]\right| \\ &= |h^q(X(n)) - h^{\tilde{q}}(X(n)) + J(q) - J(\tilde{q})|. \end{aligned} \tag{8.7}$$

Since the quantities $E_x^q[\sum_0^{\tau-1} r(X(t))]$ and $E_x^q[\tau]$ are *independent* of the value q for $x \neq L$ in $\mathbb{N}$ and $h^q(L) = 0$ for all $0 \leq q \leq 1$, it is now plain from (5.4) that the relation

$$|h^q(x) - h^{\tilde{q}}(x)| = E_x^q[\tau]|J(q) - J(\tilde{q})| \tag{8.8}$$

holds for $0 \leq q, \tilde{q} \leq 1$ and x in $\mathbb{N}$. The mapping $q \to J(q) = \mathbb{E}^q r(X)$ given by (5.3) is monotone increasing and continuously differentiable owing to Lemma 5.4. The relation (8.3) readily follows from (8.7)-(8.8) and from the linear growth of $x \to E_x^q[\tau]$. □

Lemma 8.2 *Whenever (3.5) holds, the sequences* $\{X(n)\}_0^\infty$, $\{r(X(n))\}_0^\infty$, $\{X(n)r(X(n))\}_0^\infty$ *and* $\{r^2(X(n))\}_0^\infty$ *are all uniformly integrable under* P^α.

Proof: Since the mapping r is non-negative, it is plain from the assumptions that the RV's Ξ, $r(\Xi)$, $\Xi r(\Xi)$ and $r^2(\Xi)$ must also be integrable. The results easily follows from (7.7) by virtue of the monotonicity of the mapping r. □

A proof of Theorem 7.1 is now presented with the help of Lemmas 8.1-8.2.

A proof of Theorem 7.1

Since $J(q) = \mathbb{E}^q r(X) = J^r(f^q)$ as a combined result of (5.3), Lemma 5.1 and Lemma 8.2, the relation (8.4) allows a rewritting of $\xi(n)$ as

$$\begin{aligned} -\xi(n) &= r(X(n+1)) - J(\eta(n)) \\ &= h^{\eta(n)}(X(n+1)) - E^{\eta(n)}[h^{\eta(n)}(X(n+2))|\mathbb{F}_{n+1}] \\ &= \xi_1(n) + \xi_2(n) + \xi_3(n) \end{aligned}$$

where

$$\xi_1(n) := h^{\eta(n)}(X(n+1)) - E^{\eta(n)}[h^{\eta(n)}(X(n+1))|\mathbb{F}_n],$$

$$\xi_2(n) := E^{\eta(n)}[h^{\eta(n)}(X(n+1))|\mathbb{F}_n] - E^{\eta(n+1)}[h^{\eta(n+1)}(X(n+2))|\mathbb{F}_{n+1}]$$

and

$$\xi_3(n) := E^{\eta(n+1)}[h^{\eta(n+1)}(X(n+2))|\mathbb{F}_{n+1}] - E^{\eta(n)}[h^{\eta(n)}(X(n+2))|\mathbb{F}_{n+1}]$$

for all n in $\mathbb{N}$. It now suffices to show that the relation (7.10) holds for each one of the sequences $\{\xi_i(n)\}_0^\infty$, $1 \le i \le 3$. Pose $A(n) := \sum_{i=0}^{n-1} a_i \xi_1(i)$ and $B(n) := \sum_{i=0}^{n-1} a_i \xi_2(i)$ for $n = 1, 2, \cdots$.

To show (7.10) for $\{\xi_1(n)\}_0^\infty$, first observe that

$$E^{\eta(n)}[h^{\eta(n)}(X(n+1))|\mathbb{F}_n] = E^{\alpha}[h^{\eta(n)}(X(n+1))|\mathbb{F}_n], \qquad n = 0, 1, \cdots \quad (8.9)$$

whence

$$E^{\alpha}[\xi_1(n)|\mathbb{F}_n] = 0. \qquad n = 0, 1, \cdots \quad (8.10)$$

and the RV's $\{A(n)\}_1^\infty$ form an $(P^\alpha, \mathbb{F}_n)$-martingale. It thus suffices to show the P^α-a.s. convergence of the martingale $\{A(n)\}_1^\infty$, in which case the sequence $\{A(n)\}_1^\infty$ would form a P^α-a.s. Cauchy sequence, a fact which readily implies (7.10). Note that for $0 \le i < j$,

$$\begin{aligned} E^{\alpha}[a_i\xi_1(i)a_j\xi_1(j)] &= E^{\alpha}[a_i\xi_1(i)E^{\alpha}[a_j\xi_1(j)|\mathbb{F}_{i+1}]] \\ &= E^{\alpha}[a_i\xi_1(i)E^{\alpha}[A(j+1) - A(j)|\mathbb{F}_{i+1}]] \\ &= E^{\alpha}[a_i\xi_1(i)[A(i+1) - A(i+1)]] = 0, \end{aligned} \quad (8.11)$$

and routine calculations easily give

$$E^{\alpha}[|A(n)|^2] = E^{\alpha}[\sum_{i=0}^{n-1} a_i^2 \xi_1^2(i)]. \qquad n = 1, 2, \cdots \quad (8.12)$$

By virtue of (8.1),

$$E^{\alpha}[\sum_{i=0}^{n-1} a_i^2 \xi_1^2(i)] \le K^2 \sum_{i=0}^{n-1} a_i^2 E^{\alpha}(1 + r(X(i)))^2 \le K^2 B_1 \sum_{i=0}^{\infty} a_i^2 < \infty \quad (8.13)$$

with $B_1 := \sup_n E^{\alpha}(1 + r(X(n)))^2 < \infty$ as a result of Lemma 8.2. Thus $\sup_n E^{\alpha}|A(n)|^2 < \infty$ by (8.12) and the sequence $\{A(n)\}_1^\infty$ is thus uniformly integrable under P^α. By the Martingale Convergence Theorem [2, Thm. 5.1, pp. 278], the martingale $\{A(n)\}_1^\infty$ is therefore *closable* in that the convergence

$$\lim_{n \uparrow \infty} A(n) = A(\infty) \qquad P^\alpha - a.s. \quad (8.14)$$

takes place, with the RV $A(\infty)$ being P^α-integrable. This proves (7.10) for $\{\xi_1(n)\}_0^\infty$.

To prove (7.10) for $\{\xi_2(n)\}_0^\infty$, note that for all $n < k$, the relation

$$B(k) - B(n) = -\sum_{i=n}^{k-1}(a_{i-1} - a_i)E^{\eta(i)}[h^{\eta(i)}(X(i+1))|\mathbb{F}_i]$$
$$+ a_{n-1}E^{\eta(n)}[h^{\eta(n)}(X(n+1))|\mathbb{F}_n] - a_{k-1}E^{\eta(k)}[h^{\eta(k)}(X(k+1))|\mathbb{F}_k]$$

holds, and implies

$$\begin{aligned}|B(k) - B(n)| \leq K\sum_{i=n}^{k-1}(a_{i-1} - a_i)(1+X(i))(1+r(X(i)))\\ + Ka_{n-1}(1+X(n))(1+r(X(n)))\\ + Ka_{k-1}(1+X(k))(1+r(X(k)))\end{aligned} \qquad P^\alpha - a.s. \quad (8.15)$$

by a direct use of (8.2). By making use of (7.6), it is easy to see that the relation (8.15) implies

$$\begin{aligned}|B(k) - B(n)| \leq K\sum_{i=n}^{\infty}(a_{i-1} - a_i)(1+X(i))(1+r(X(i)))\\ + \tilde{K}a_{n-1}(1+X(n))(1+r(X(n)))\end{aligned} \qquad P^\alpha - a.s. \quad (8.16)$$

for some positive constant $\tilde{K}$. For each $n \geq 1$ in $\mathbb{N}$, pose

$$\Lambda(n) := K\sum_{i=n}^{\infty}(a_{i-1} - a_i)(1+X(i))(1+r(X(i))) \qquad (8.17)$$

and

$$e(n) := \tilde{K}a_{n-1}(1+X(n))(1+r(X(n))). \qquad (8.18)$$

With this notation, (8.16) becomes for all $k > n$,

$$|B(k) - B(n)| \leq \Lambda(n) + e(n). \qquad P^\alpha - a.s. \quad (8.19)$$

The estimate (7.7) again yields the bound

$$(1+X(n))(1+r(X(n))) \leq B_L + \Xi + r(\Xi) + \Xi r(\Xi), \qquad P^\alpha - a.s. \quad (8.20)$$

with $B_L := 1 + (L+1) + r(L+1) + (L+1)r(L+1)$, and this implies the convergence

$$\lim_{n\uparrow\infty} e(n) = \lim_{n\uparrow\infty} a_{n-1}(1+X(n))(1+r(X(n))) = 0 \tag{8.21}$$

both P^α-a.s. and in $L^1(\Omega, \mathcal{F}, P^\alpha)$ (since the right-hand side of (8.20) is integrable by (3.5)). An easy application of Lemma 8.2 again gives

$$E^\alpha \Lambda(n) = KE^\alpha[\sum_{i=n}^{\infty}(a_{i-1}-a_i)(1+X(i))(1+r(X(i)))] \leq KB_2 \sum_{i=n}^{\infty}(a_{i-1}-a_i) < \infty,$$

with $B_2 := \sup_n E^\alpha(1+X(n))(1+r(X(n))) < \infty$, and the convergence

$$\lim_{n\uparrow\infty} \Lambda(n) = 0 \tag{8.22}$$

takes place in $L^1(\Omega, \mathcal{F}, P^\alpha)$, thus in probability. The sequence $\{\Lambda(n)\}_0^\infty$ being positive and monotone-decreasing, convergence in probability (to zero) of the sequence $\{\Lambda(n)\}_1^\infty$ implies the a.s. convergence (to zero) under P^α. Since

$$\sup_{k>n} |\sum_{n}^{k-1} a_i \xi_2(i)| = \sup_{k>n} |B(k) - B(n)| \leq \Lambda(n) + e(n)$$

owing to (8.19), the convergence

$$\lim_{n\uparrow\infty} (\sup_{k\geq n} |\sum_{i=n}^{k} a_i \xi_2(i)|) = 0 \qquad P^\alpha - a.s. \tag{8.23}$$

readily takes place, and this proves (7.10) for the sequence $\{\xi_2(n)\}_0^\infty$.

To prove (7.10) for $\{\xi_3(n)\}_0^\infty$, note from (8.3) that for all $n \geq 1$,

$$|\xi_3(n)| \leq K(1+X(n+1))|\eta(n+1) - \eta(n)|, \tag{8.24}$$

whereas the recursion (3.3) implies

$$|\eta(n+1) - \eta(n)| \leq a_n |V - r(X(n+1))|.$$

Consequently, the relation

$$|\xi_3(n)| \leq K(V+1)a_n(1+X(n+1))(1+r(X(n+1))) \tag{8.25}$$

holds, and (8.21) again implies

$$\lim_{n\uparrow\infty} |\xi_3(n)| = 0 \qquad P^\alpha - a.s. (8.26)$$

or equivalently,

$$\lim_{n\uparrow\infty} \sup_{n\le k\le m(n,T)} |\xi_3(k)| = 0 \qquad P^\alpha - a.s. (8.27)$$

for every $T > 0$. From the fact $\sum_{i=n}^{m(n,T)} a_i \le T$, it now follows that

$$\sup_{n\le k\le m(n,T)} |\sum_{i=n}^{k} a_i \xi_3(i)| \le \sum_{i=n}^{m(n,T)} a_i |\xi_3(i)| \le T(\sup_{n\le k\le m(n,T)} |\xi_3(k)|)$$

and the convergence

$$\lim_{n\uparrow\infty} (\sup_{n\le k\le m(n,T)} |\sum_{i=n}^{k} a_i \xi_3(i)|) = 0 \qquad P^\alpha - a.s. (8.28)$$

immediately obtains from (8.27). This completes the proof of Theorem 7.1. □

9. Some remarks

At this point, the reader may wonder whether knowledge of the threshold value L can be relaxed and whether a similar Stochastic Approximations algorithm can be used to generate a sequence of estimates for the threshold value L? One possible approach would consist in generating the estimates $\{L(n)\}_0^\infty$ for the threshold value L via a scheme which parallels to (3.3): First generate a sequence $\{\Lambda(n)\}_0^\infty$ of $\mathbb{R}_+$-valued RV's through the recursion

$$\Lambda(n+1) = \left[\Lambda(n) + a_n(V - r(X(n+1)))\right]^+ \qquad n = 0, 1, \cdots (9.1)$$

with $\Lambda(0) \ge 0$, where the notation $[x]^+ = 0 \vee x$ is used for all x in $\mathbb{R}$. The estimate $L(n)$ is then given by $L(n) := \lfloor \Lambda(n) \rfloor$, i.e., the so-called integer part of $\Lambda(n)$. An adaptive policy α can then be generated accordingly via the Certainty Equivalence principle. The threshold control being integer-valued, whenever the RV $\Lambda(n)$ moves across the boundary of an integer, a certain amount of discontinuity incurs on the corresponding transition probabilities. In constrast with the case studied in previous sections, this fact prevents from establishing some Lipschitz properties associated with the RV's $\{\Lambda(n)\}_0^\infty$, as the ones given in Lemma 8.1, and

the argument presented in Sections 7-8 thus collapes in this case. This discontinuity also precludes a direct use of weak convergence results [4].

This points out to a class of situations where on-line estimates for integer-valued quantities need to be generated. New techniques seem required to handle such problems, and research along these lines would be of interest from both theoretical and practical standpoints since integer-valued controls often occur in the control of queues.

REFERENCES

[1] M. Gerla and L. Kleinrock, "Flow control: A comparative survey," *IEEE Trans. Commun.*, vol. COM-28, pp. 553-574 (1980).

[2] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, Second Edition, Academic Press, New York, 1975.

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

[4] H. J. Kushner and A. Shwartz, "An invariant measure approach to the convergence of Stochastic Approximations with state-dependent noise," *SIAM J. Control Opt.*, Vol. 22, pp.13-27 (1984).

[5] D.-J. Ma and A. M. Makowski, "A simple problem of flow control I: Optimality Results," *IEEE Trans. Auto. Control*, submitted (1987).

[6] D.-J. Ma and A. M. Makowski, "Parameter estimation, identification and adaptive implementation of threshold policies for a system of flow control," in preparation (1987).

[7] D.-J. Ma, A. M. Makowski and A. Shwartz, "Estimation and optimal control for constrained Markov chains," *Proceedings of the 25th IEEE Conference on Decision and Control*, Athens, Greece, pp. 994-999 (1986).

[8] D.-J. Ma, A. M. Makowski and A. Shwartz, "Stochastic Approximation for constrained Markov decision processes," *Proceedings of the 1987 International Symposium on the Mathematical Theory of Networks and Systems*, Phoenix, Arizona (June 1987).

[9] A. M. Makowski and A. Shwartz, "Implementation issues for Markov decision processes," *Proceedings of a Workshop on the Stochastic Differential Systems*, Institute of Mathematics and its Applications, University of Minnesota, Eds. W. Fleming and P.-L. Lions, Springer Verlag Lecture Notes in Control and Information Sciences (1986).

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

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

[12] A. Shwartz and A. M. Makowski, "Adaptive policies for a system of competing queues I: Convergence results for the long-run average cost", *Adv. Appl. Prob.*, submitted (1987).