Proceedings of the 33rd
Conference on Decision and Control
Lake Buena Vista, FL - December 1994

DENUMERABLE CONTROLLED MARKOV CHAINS WITH STRONG AVERAGE OPTIMALITY CRITERION: BOUNDED & UNBOUNDED COSTS

Rolando Cavazos-Cadena
Departamento de Estadística y Cálculo
UAAAN
Buenavista 25315
Saltillo, COAH
MEXICO

Emmanuel Fernández-Gaucherand
Systems & Industrial Eng. Dept.
The University of Arizona
Tucson, AZ 85721
USA
Email: emmanuel@sie.arizona.edu.

1. Introduction

We study discrete-time controlled stochastic systems, modeled by controlled Makov chains (CMC) with denumerable state space and compact actions space, and with an infinite planning horizon. One optimality criterion often used is the long-run expected average cost (AC); see [1]. A classical approach to study AC optimality consists in formulating the AC case as a limit of the discounted cost (DC) case, as the discount factor increases to 1, i.e., as the discounting effect *vanishes.* This approach has been rekindled in recent years, with the introduction by Sennott and others of conditions under which AC optimal stationary policies are shown to exist [4]-[5], [9], [16], [19], [20]-[22], [18]; see also [1], [7]. However, the AC criterion is a rather underselective criterion, since it completely neglects the finite-time evolution of the state/cost process [1], [11]. Therefore, it is of much theoretical and practical interest to obtain conditions under which stonger results than AC optimality can be obtained.

Our main interest in this paper is to study the relation between the notions of AC optimality and *strong* average cost (SAC) optimality. The latter criterion is introduced to asses the performance of a policy over long but finite horizons, as well as in the long-run average sense. We show that for bounded one-stage cost functions, Sennott's conditions are sufficient to guarantee that *every* AC optimal policy is also SAC optimal. On the other hand, a detailed counterexample is given that shows that the latter result does not extend to the case of unbounded cost functions.

The paper is organized as follows. In section 2 the model is presented. Section 3 defines the stochastic control problem, with an expected average cost criterion. Section 4 introduces the strong average optimality criterion. In section 5, the case of bounded cost functions is studied, while unbounded costs are treated in section 6. A complete version of this paper may be obtained from the second author.

2. The Model

We study controlled Markov chains (CMC) described by the four-tuples $\langle \mathbf{S}, \mathbf{U}, \mathcal{U}, p \rangle$, where the state space $\mathbf{S}$ is a denumerable set, endowed with the discrete topology; the metric space $\mathbf{U}$ denotes the control or action set. Furthermore, for each $x \in \mathbf{S}$, a nonempty compact set $\mathcal{U}(x) \subset \mathbf{U}$ specifies the admissible actions when the system is in state x. Let $\mathbf{K} := \{(x,u) | x \in \mathbf{S}, u \in \mathcal{U}(x)\}$ denote the space of admissible state-action pairs, which is considered as a topological subspace of $\mathbf{S} \times \mathbf{U}$. The state transition law is given by $p : (x,y,u) \mapsto p_{x,y}(u)$, an stochastic kernel on $\mathbf{S}$ given $\mathbf{K}$; see [1], [15].

In addition, to assess the performance of the system, a measurable† (and possibly unbounded) one-stage cost function $c : \mathbf{K} \to \mathbb{R}$ is chosen by the decision-maker. Thus, at time $t \in \mathbb{N}_0 := \{0,1,2,\ldots\}$, the system is observed to be in some state, say $X_t = x \in \mathbf{S}$, and a control/decision $U_t = u \in \mathcal{U}(x)$ is taken. Then a cost $c(x,u)$ is incurred, and by the next control/decision epoch $t+1$, the state of the system will have evolved to $X_{t+1} = y$ with probability $p_{x,y}(u)$. The following is an standard assumption.

Assumption 2.1:
(i) For each x, $y \in \mathbf{S}$, the mappings $u \mapsto c(x,u)$ and $u \mapsto p_{x,y}(u)$ are continuous in $u \in \mathcal{U}(x)$.
(ii) The cost function $c : \mathbf{K} \to \mathbb{R}$ is bounded below. □

The available information for control at time $t \in \mathbb{N}_0$ is given by the *history* of the process up to that time $H_t := (X_0, U_0, X_1, U_1, \ldots, U_{t-1}, X_t)$, which is a random process taking values in the *history spaces* $\mathbf{H}_t$, given by

$$\mathbf{H}_0 := \mathbf{S}, \qquad \mathbf{H}_t := \mathbf{H}_{t-1} \times \mathbf{K},$$

and the canonical sample space is given by $\Omega_\infty := (\mathbf{S} \times \mathbf{U})^\infty$; see [1].

† Given a topological space $\mathbf{W}$, its Borel σ-algebra will be denoted by $\mathcal{B}(\mathbf{W})$; measurability will be always understood as Borel measurability henceforth.

An *admissible control policy* is a (possibly randomized) rule for choosing actions, which may depend on the entire history of the process H_t up to the present time; see [1] for an extensive discussion. Thus, a policy is especified by a sequence $\pi = \{\pi_t\}_{t\in \mathbb{N}_0}$ of stochastic kernels π_t on $\mathbf{U}$ given $\mathbf{H}_t$, such that for each $h_t \in \mathbf{H}_t$, $\pi_t(\cdot \mid h_t)$ is a probability measure on $\mathcal{B}(\mathbf{U})$, concentrated on $\mathcal{U}(x)$. The set of all admissible policies will be denoted by $\mathbf{\Pi}$. In the sequel, the class of *stationary* (deterministic) policies will be of particular interest. A policy $\pi \in \mathbf{\Pi}$ is said to be stationary (deterministic) if there exists a control/decision function $f : \mathbf{S} \to \mathbf{U}$, such that $U_t = f(x) \in \mathcal{U}(x)$ is the action prescribed by π at time t, if $X_t = x$. The set of all stationary deterministic policies is denoted as $\mathbf{\Pi}_{SD}$.

Given the initial state $X_0 = x \in \mathbf{S}$, and a policy $\pi \in \mathbf{\Pi}$, the corresponding state/action and history processes, $\{X_t, U_t\}$ and $\{H_t\}$ respectively, are random processes defined on the canonical probability space $(\mathbf{\Omega}_\infty, \mathcal{B}(\mathbf{\Omega}_\infty), \mathcal{P}_x^\pi)$ via the projections $X_t(h_\infty) := x_t$, $U_t(h_\infty) := u_t$ and $H_t(h_\infty) := h_t$, for each $h_\infty = (x, u_0, \ldots, x_t, u_t, \ldots) \in \mathbf{\Omega}_\infty$, where $\mathcal{P}_x^\pi$ is uniquely determined; see[1], [3], [15]. The corresponding expectation operator is denoted by $\mathbb{E}_x^\pi$.

3. Stochastic Optimal Control

Our main interest in this paper is to study the relation between the notions of AC optimality and *strong* average cost (SAC) optimality. The latter criterion is introduced to asses the performance of a policy over long but finite horizons, as well as in the long-run average sense. The standard approach [1] to study the AC stochastic control problem as a limit of the dicounted cost (DC) problem will be followed. The criteria that will be used in subsequent developments are given below.

Discounted Cost (DC): For a discount factor $0 < \alpha < 1$, the DC incurred by $\pi \in \mathbf{\Pi}$, when the initial state of the system is $x \in \mathbf{S}$, is given by

$$V_\alpha(x, \pi) := \lim_{n\to\infty} \mathbb{E}_x^\pi \left[\sum_{t=0}^{n} \alpha^t c(X_t, U_t) \right],$$

and the optimal α-discounted *value function* is defined as

$$V_\alpha^*(x) := \inf_{\pi\in\mathbf{\Pi}} \{ V_\alpha(x, \pi) \}. \tag{3.1}$$

A policy $\pi \in \mathbf{\Pi}$ is said to be DC optimal, for the discount factor α, if $V_\alpha(x, \pi) = V_\alpha^*(x)$, for all $x \in \mathbf{S}$.

Under Assumption 3.1 below, $V_\alpha^*(\cdot)$ satisfies *Bellman's Optimality Equation* (also called the discounted cost optimality equation (DCOE)), i.e.,

$$V_\alpha^*(x) = \inf_{u\in\mathcal{U}(x)} \{ c(x, u) + \alpha \sum_{y\in\mathbf{S}} p_{x,y}(u) V_\alpha^*(y) \}, \forall x \in \mathbf{S}; \tag{3.2}$$

see [1], [2], [3], [15].

Assumption 3.1: For each $x \in \mathbf{S}$, $V_\alpha^*(x) < \infty$. □

Total Expected Cost Over Finite Horizons (FHC): The total expected cost incurred by the policy $\pi \in \mathbf{\Pi}$ over a planning horizon of $n \in \mathbb{N}$ stages, when the initial state of the system is $x \in \mathbf{S}$, is given by

$$V_n(x, \pi) := \mathbb{E}_x^\pi \left[\sum_{t=0}^{n-1} c(X_t, U_t) \right], \tag{3.3}$$

and the optimal n-horizon *value function* is defined as

$$V_n^*(x) := \inf_{\pi\in\mathbf{\Pi}} \{ V_n(x, \pi) \}. \tag{3.4}$$

From Assumptions 2.1 and 3.1 if follows that $V_n^*(x)$ is finite for each $x \in \mathbf{S}$. Moreover, there exists an n-horizon optimal policy $\pi_n^* \in \mathbf{\Pi}$, i.e.,

$$V_n(x, \pi_n^*) = V_n^*(x), \tag{3.5}$$

see [1], [2], [3], [15].

Average Cost: The long-run expected average cost obtained by $\pi \in \mathbf{\Pi}$, when the initial state of the system is $x \in \mathbf{S}$, is given by

$$J(x, \pi) := \limsup_{n\to\infty} \frac{1}{n} \mathbb{E}_x^\pi \Big[\sum_{t=0}^{n-1} c(X_t, U_t) \Big], \tag{3.6}$$

and the optimal expected average cost is defined as

$$J^*(x) := \inf_{\pi\in\mathbf{\Pi}} \{ J(x, \pi) \}. \tag{3.7}$$

A policy $\pi \in \mathbf{\Pi}$ is said to be AC optimal if $J(x, \pi) = J^*(x)$, for all $x \in \mathbf{S}$

A classical approach to study average optimality consists in formulating the AC case as a limit of the DC case, as $\alpha \uparrow 1$, i.e., as the discounting effect *vanishes*; see [1] and references therein for a comprehensive discussion on the subject. This approach has been rekindled in recent years, with the introduction by Sennott and others of conditions under which AC optimal stationary policies are shown to exist [16], [14], [20]-[22], [19]; see also [1], [9]. Our analysis will be carried out under the set of asumptions introduced by Sennott [20]-[22], which are minimal with respect to several other competing assumptions; see [9]. In particular, the assumptions in [13] imply those in [20], [21]. Hence, in addition to Assumptions 2.1 and 3.1, the following assumptions will be used in the sequel.

Assumption 3.2:

(i) There exists $z^* \in \mathbf{S}$, $\beta \in (0,1)$, and $N \in [0,\infty)$ such that for all $\alpha \in (\beta, 1)$,

$$h_\alpha(x) := V_\alpha^*(x) - V_\alpha^*(z) \geq -N, \quad \forall x \in \mathbf{S}. \quad (3.8)$$

(ii) There exists a function $b : \mathbf{S} \to [0,\infty)$ such that $h_\alpha(\cdot) \leq b(\cdot)$, for all $\alpha \in (\beta, 1)$, and furthermore $\sum_y p_{x,y}(u)b(y) < \infty$, for some $(x,u) \in \mathbf{K}$. □

The main results deriving from the above assumption are summarized below; for a proof see [1], [20], [21], [7]

Lemma 3.1: Let

$$\rho_\alpha := (1-\alpha)V_\alpha^*(z), \quad (3.9)$$

for $\alpha \in (0,1)$. Then under Assumptions 2.1, 3.1-3.2, the following holds.

(i) There exists $\rho^* \in \mathbb{R}$ such that

$$\rho^* = J^*(x), \quad \forall x \in \mathbf{S}.$$

(ii) Moreover,

$$\lim_{\alpha \uparrow 1} \rho_\alpha = \rho^*.$$

(iii) There exists $h : \mathbf{S} \to \mathbb{R}$, with $-N \leq h(\cdot)$, $h(\cdot) \leq b(\cdot)$ for all $x \in \mathbf{S}$, such that the AC *Optimality Inequality* (ACOI) holds:

$$\rho^* + h(x) \geq \inf_{u \in \mathcal{U}(x)} \left[c(x,u) + \sum_{y \in \mathbf{S}} p_{x,y}(u)h(y) \right]. \quad (3.10)$$

(iv) For each $x \in \mathbf{S}$, the term within brackets in (3.10) is a lower semi-continuous function of $u \in \mathcal{U}(x)$, and thus it has a minimizer $f^*(x) \in \mathcal{U}(x)$. Moreover, any policy $f^* \in \mathbf{\Pi}_{SD}$ attaining the minimum in the ACOI is AC optimal.

4. Strong Average Optimality Criterion

Definition 4.1: A policy $\pi^* \in \mathbf{\Pi}$ is said to be *strong average cost* (SAC) optimal if

$$\frac{1}{n+1}\left[V_n(x,\pi^*) - V_n^*(x)\right] \underset{n\to\infty}{\longrightarrow} 0, \quad \forall x \in \mathbf{S}. \quad (4.1)$$

Thus, a policy $\pi^* \in \mathbf{\Pi}$ is SAC optimal if the difference between the average cost for horizon n incurred under π^* and the optimal average cost for horizon n vanishes as $n \to \infty$. This property thus ensures that π^* is a policy inducing good performance for long but finite horizons, which is indeed a very desirable property to look for in infinite horizon average optimal policies. Notice that every policy that is SAC optimal is also AC optimal, however the opposite is not necessarily true; see [12], [13]. Therefore, in the sequel we study the following.

Question: *Is an arbitrary AC optimal policy also SAC optimal under Assumptions 2.1 and 3.1-3.2?*

We will show that the answer to the above question is affirmative, when the one-stage cost function is bounded. However, for unbounded cost functions, a counterexample is provided which shows that the answer is negative, in general, and in particular for a policy as in Lemma 3.1(iv). A related study to the question above is [13], where SAC optimality was shown for every AC optimal policy, under conditions stronger than those used in the sequel.

5. Bounded Costs Case

The following result is similar to results in [13], but here it is proved under a different set of assumptions; this result will be fundamental to prove the main result in this section.

Lemma 5.1: Suppose that the one-stage cost function $c(\cdot,\cdot)$ is (uniformly) bounded in $\mathbf{K}$. Then, under Assumptions 2.1, 3.1-3.2, the following holds:

$$\frac{V_n^*(x)}{n+1} \underset{n\to\infty}{\longrightarrow} \rho^*, \quad \forall x \in \mathbf{S}. \quad (5.1)$$

Proof: Let $\beta \in (0,1)$ be as in Assumption 3.2, and pick $\alpha \in (\beta, 1)$. The DCOE (3.2) can equivalently be written as (see [1], [15]),

$$\rho_\alpha + h_\alpha(x) = \inf_{u \in \mathcal{U}(x)} \left\{ c(x,u) + \alpha \sum_{y \in \mathbf{S}} p_{x,y}(u)h_\alpha(y) \right\},$$

$\forall x \in \mathbf{S}$, where $h_\alpha(\cdot)$ and ρ_α are as in (3.8) and (3.9), respectively. Then, for every $(x,u) \in \mathbf{K}$ it follows that

$$\rho_\alpha + h_\alpha(x) \leq c(x,u) + \alpha \sum_{y \in \mathbf{S}} p_{x,y}(u)h_\alpha(y). \quad (5.2)$$

Since $c(\cdot,\cdot)$ is bounded, then

$$-N \leq h_\alpha(x) \leq \frac{2\|c\|}{1-\alpha} < \infty,$$

where $\|c\| := max\{|c(x,u)| \,|\, (x,u) \in \mathbf{K}\}$, which is finite by the boundedness assumption. Therefore, defining $\tilde{h}_\alpha(\cdot) := h_\alpha(\cdot) + N$, with N as in Assumption 3.2, it follows that

$$0 \leq \tilde{h}_\alpha(x) \leq \frac{2\|c\|}{1-\alpha} + N < \infty. \quad (5.3)$$

Moreover, it is not difficult to see that (5.2) is equivalent to the following

$$\rho_\alpha - N(1-\alpha) + \tilde{h}_\alpha(x) \le c(x,u) + \alpha \sum_{y\in\mathbf{S}} p_{x,y}(u)\tilde{h}_\alpha(y),$$

and since $\tilde{h}(\cdot) \ge 0$, this last inequality implies that

$$\rho_\alpha - N(1-\alpha) + \tilde{h}_\alpha(x) \le c(x,u) + \sum_{y\in\mathbf{S}} p_{x,y}(u)\tilde{h}_\alpha(y),$$

$\forall (x,u) \in \mathbf{K}$. Then, using standard arguments (see [1], [15], [23], [20], [21]), it follows that

$$\rho_\alpha - N(1-\alpha) + \frac{\tilde{h}_\alpha(x)}{n+1} \le \frac{\mathbb{E}_x^\pi\left[\sum_{t=0}^n c(X_t,U_t)\right]}{n+1} + \frac{\mathbb{E}_x^\pi\left[\tilde{h}_\alpha(X_{n+1})\right]}{n+1},$$

for all $x \in \mathbf{S}$, $\pi \in \mathbf{\Pi}$, and $n \in \mathbb{N}$. Using the policy π_n^* in (3.5), (5.3) and the above inequality yield

$$\rho_\alpha - N(1-\alpha) + \frac{\tilde{h}_\alpha(x)}{n+1} \le \frac{V_n^*(x)}{n+1} + \frac{2\|c\|}{(1-\alpha)(n+1)} + \frac{N}{n+1},$$

and taking the limit inferior in both sides as $n \to \infty$, it follows that

$$\rho_\alpha - N(1-\alpha) \le \liminf_{n\to\infty} \frac{V_n^*(x)}{n+1},$$

and then letting $\alpha \uparrow 1$, Lemma 3.1(ii) yields that

$$\rho^* \le \liminf_{n\to\infty} \frac{V_n^*(x)}{n+1}. \tag{5.4}$$

Now let $f^* \in \mathbf{\Pi}_{SD}$ be the policy in Lemma 3.1(iv) then, $\forall x \in \mathbf{S}$,

$$\rho^* + h(x) \ge c(x, f^*(x)) + \sum_{y\in\mathbf{S}} p_{x,y}(f^*(x))h(y).$$

A simple induction argument then gives that

$$\rho^* + \frac{h(x)}{n+1} \ge \frac{\mathbb{E}_x^{f^*}\left[\sum_{t=0}^n c(X_t,U_t)\right]}{n+1} + \frac{\mathbb{E}_x^{f^*}\left[h(X_{n+1})\right]}{n+1}.$$

From the above, together with (3.3), (3.4) and (3.8), it follows that

$$\rho^* + \frac{h(x)}{n+1} \ge \frac{V_n^*(x)}{n+1} - \frac{N}{n+1}, \quad \forall x \in \mathbf{S}, n \in \mathbb{N}.$$

Therefore, one obtains that

$$\rho^* \ge \limsup_{n\to\infty} \frac{V_n^*(x)}{n+1}. \tag{5.5}$$

Thus, the result follows by combining (5.4) and (5.5). □

The following is the main result of this section, which gives a positive partial answer to the question posed in Section 4.

Theorem 5.1: Suppose that the one-stage cost function $c(\cdot,\cdot)$ is (uniformly) bounded in $\mathbf{K}$. Then, under Assumptions 2.1, 3.1-3.2, *every* AC optimal policy is SAC optimal.

Proof: Let $\pi^* \in \mathbf{\Pi}$ be *any* average optimal policy. Hence, for each $x \in \mathbf{S}$,

$$\limsup_{n\to\infty} \frac{V_n(x,\pi^*)}{n+1} = J^*(x) = \rho^*, \tag{5.6}$$

by Lemma 3.1(i). On the other hand $V_n(x,\pi^*) \ge V_n^*(x)$, and thus

$$\liminf_{n\to\infty} \frac{V_n(x,\pi^*)}{n+1} \ge \liminf_{n\to\infty} \frac{V_n^*(x)}{n+1} = \rho^*,$$

by Lemma 5.1. The last inequality and (5.6) combined give that

$$\frac{V_n(x,\pi^*)}{n+1} \underset{n\to\infty}{\longrightarrow} \rho^*, \quad \forall x \in \mathbf{S}. \tag{5.7}$$

Finally, from (5.1) and (5.7) it follows that

$$\frac{V_n(x,\pi^*) - V_n^*(x)}{n+1} \underset{n\to\infty}{\longrightarrow} (\rho^* - \rho^*) = 0.$$

Thus π^* is strong average optimal. □

6. A Counterexample for the Unbounded Costs Case

In this section an example is given showing that the result in Theorem 5.1 does *not* extend to the case of unbounded one-stage cost functions. In particular, a policy as in Lemma 3.1(iv) is exhibited, which is not SAC optimal. This example exploits the fact that Assumptions 3.1-3.2 involve the one-stage cost and the probabilistic structure of the controlled Markov chain only indirectly, i.e., through the derived quantities $V_\alpha^*(\cdot)$, but no explicit conditions are given on the primary model parameters $c(\cdot,\cdot)$ and $[p_{x,y}(u)]$. In the literature [20]-[22], [4]-[6], [24], explicit conditions have been imposed on $c(\cdot,\cdot)$ and $[p_{x,y}(u)]$ in order to verify Assumptions 3.1-3.2; see also [9]. In [13] explicit conditions, which imply Assumptions 3.1-3.2 [9], were imposed on $c(\cdot,\cdot)$ and $[p_{x,y}(u)]$ to show SAC optimality of AC optimal policies.

Example 6.1: Consider a CMC with state space $\mathbf{S} = \mathbb{N}_0$. To specify the other components of the model, let $\beta \in (0,1)$ be *fixed*, and select a sequence $\{t_k\} \subset \mathbb{N}_0$ such that
(a) $0 = t_0 < t_1 < t_2 < \ldots$, and
(b)

$$t_k > k\Big[\sum_{s=0}^{k-1} \frac{(1+t_s)}{\beta^{t_s}}\Big], \qquad k \in \mathbb{N}.$$

Next, set $\mathbf{U} = \{0,1\}$ (endowed with the discrete topology) and define the action set, the cost function and the transition law as follows:
(i) for $x \neq t_k$, $k \in \mathbb{N}$, let $\mathcal{U}(x) := \{1\}$, $c(x,1) := 0$, and $p_{x,x+1}(1) := 1$;
whereas
(ii) for $x = t_k$, $k \in \mathbb{N}$, let $\mathcal{U}(t_k) := \{0,1\}$, $c(t_k,0) = c(t_k,1) := \frac{(1+t_k)}{\beta^{t_k}}$, and $p_{t_k,0}(0) := 1$, $p_{t_k,1+t_k}(1) = 1$.

Note that $c(0,u) = 1$, for $u = 0,1$. Thus, in state $X_t = x \neq t_k$, $k \in \mathbb{N}$, the only available action is $U_t = 1$, which produces a state transition to $X_{t+1} = x+1$, at no cost. On the other hand, when $X_t = t_k$, for some $k \in \mathbb{N}$, both actions are available: $U_t = 0$ produces a state transition to $X_{t+1} = 0$, and $U_t = 1$ to $X_{t+1} = 1 + t_k$; in either case the cost incurred is $c(t_k, U_t) = \frac{(1+t_k)}{\beta^{t_k}}$.

In the example above, Assumption 2.1 clearly holds. Furthermore, we have shown that Assumptions 3.1-3.2 are also satisfied. Let f_0 and f_1 be stationary policies given by

$$f_0(t_k) = 0; \qquad f_1(t_k) = 1, \quad k \in \mathbb{N}. \tag{6.1}$$

Thus, under action of policy f_0, the state $z^* = 0$ is absorbing, and it is clear that

$$V_\alpha(0, f_0) = \frac{1}{1-\alpha}, \tag{6.2}$$

and, for $k \in \mathbb{N}$,

$$\begin{aligned} V_\alpha(t_k, f_0) &= c(t_k, 0) + \alpha V_\alpha(0, f_0) \\ &= \frac{1+t_k}{\beta^{t_k}} + \frac{\alpha}{1-\alpha}. \end{aligned} \tag{6.3}$$

More generally, if the initial state $x \in \mathbb{N}_0$ is such that $t_k < x < t_{k+1}$, then under the action of policy f_0 no cost will be incurred until state t_{k+1} is reached, which occurs after $(t_{k+1} - x)$ time periods. Therefore, one obtains that for $t_k < x < t_{k+1}$, $k \in \mathbb{N}$,

$$V_\alpha(x, f_0) = \alpha^{(t_{k+1}-x)} V_\alpha(t_{k+1}, f_0). \tag{6.4}$$

Then, the following can be shown.

Proposition 6.1: For each $\alpha \in (\beta, 1)$, the policy f_0 in (6.1) is DC optimal for the CMC in Example 6.1.

Proposition 6.2: For the CMC in Example 6.1, Assumptions 3.1-3.2 are satisfied. In addition, the policy f_0 in (6.1) attains the minimum in the ACOI, and consequently is AC optimal.

Proposition 6.3: For the CMC in Example 6.1, and the policy f_0 in (6.1), the following holds:

$$\limsup_{n\to\infty} \frac{V_n(0, f_0) - V_n^*(0)}{n+1} = 1. \tag{6.5}$$

Therefore, f_0 is *not* SAC optimal.

In summary, we have shown that the CMC in Example 6.1 satisfies Assumptions 2.1, 3.1 and 3.2, but the AC optimal policy f_0 is *not* SAC optimal. Furthermore, it is not difficult to verify that, for all $x \in \mathbb{N}_0$,

$$\frac{V_n(x, f_0)}{n+1} \xrightarrow[n\to\infty]{} 1,$$

and that

$$\liminf_{n\to\infty} \frac{V_n(x, f_1)}{n+1} = 0, \tag{6.6}$$

and thus

$$\limsup_{n\to\infty} \frac{V_n(x, f_0) - V_n^*(x)}{n+1} = 1.$$

In addition, (6.6) shows that using the limit inferior or superior in the definition of the AC criterion does *not* lead to equivalent criteria, answering in the negative a question posed in [7].

ACKNOWLEDGEMENTS

Research supported by a U.S.-México Collaborative Research Program funded by the National Science Foundation under grant NSF-INT 9201430, and by CONACyT-MEXICO. The first author was also partially supported by the MAXTOR Foundation for applied Probability and Statistics, under grant No. 01-01-56/04-93. The second author was partially supported by the Engineering Foundation under grant RI-A-93-10, and by a grant from the AT&T Foundation.

References

[1] A. Arapostathis, V.S. Borkar, E. Fernández Gaucherand, M.K. Ghosh and S.I. Marcus (1993) Discrete-Time controlled Markov processes with an average cost criterion: a survey. SIAM J. Control Optim. 31:282-344

[2] D.P. Bertsekas (1987) *Dynamic Programming: Deterministic and Stochastic Models.* Prentice-Hall, Englewood Cliffs.

[3] D.P. Bertsekas and S.E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.

[4] V.S. Borkar (1983) Controlled Markov chains and stochastic networks. SIAM J. Control Optim. 21:652-666.

[5] V.S. Borkar On minimum cost per unit time control of Markov chains. SIAM J. Control Optim. 22:965-984.

[6] V.S. Borkar (1991) *Topics in Controlled Markov Chains.* Pitman Research Notes in Mathematics Series #240, Longman Scientific & Technical, UK.

[7] R. Cavazos-Cadena (1991) Recent results on conditions for the existence of average optimal stationary policies. Annals Operat. Res. 28:3-28.

[8] R. Cavazos-Cadena (1991) A counterexample on the optimality equation in Markov decision chains with the average cost criterion. Syst. Control Lett. 16:387-392.

[9] R. Cavazos-Cadena and L.I. Sennott (1992) Comparing recent assumptions for the existence of average optimal stationary policies. Operat. Res. Lett. 26:33-37.

[10] E.B. Dynkin and A.A. Yushkevich (1979) *Controlled Markov Processes.* Springer-Verlag, New York.

[11] E. Fernández-Gaucherand, M.K. Ghosh and S.I. Marcus (1994) Controlled Markov processes on the infinite planning horizon: weighted and overtaking cost criteria. To appear in **ZOR**: Methods and Models of Operations Research 39:173-202.

[12] J. Flynn (1980) On optimality criteria for dynamic programs with long finite horizon. J. Math. Anal. Appl. 76:202-208.

[13] M.K. Ghosh and S.I. Marcus (1992) On strong average optimality of Markov decision processes with unbounded costs. Operat. Res. Lett. 11:99-104.

[14] D. Gatarek and L. Stettner (1990) On the compactness method in general ergodic impulsive control of Markov processes. Stoch. and Stoch. Reports 31:15-25.

[15] O. Hernández-Lerma (1989) *Adaptive Markov Control Processes.* Springer-Verlag, New York.

[16] O. Hernández-Lerma and J.B. Lasserre (1990) Average cost optimal policies for Markov control processes with Borel state space and unbounded costs. Syst. Control Lett. 15:349–356.

[18] R.K. Ritt and L.I. Sennott (1992) Optimal stationary policies in general state space Markov decision chains with finite action sets. Math. Operat. Res. 17:901-909.

[19] M. Schäl (1993) Average optimality in dynamic programming with general state space. Math. Operat. Res. 18:163-172.

[20] L.I. Sennott (1986) A new condition for the existence of optimal stationary policies in average cost Markov decision processes. Oper. Res. Lett. 5:17–23.

[21] L.I. Sennott (1989) Average cost optimal stationary policies in infinite state Markov decision processes with unbounded costs. Oper. Res. 37:626–633.

[22] L.I. Sennott (1989) Average cost semi-Markov decision processes and the control of queueing systems. Probab. in Eng. & Info. Sci. 3:247–272.

[23] S.M. Ross (1983) *Introduction to Stochastic Dynamic Programming.* Academic Press, New York.

[24] R.R. Weber and S. Stidham, Jr. (1987) Optimal control of service rates in networks of queues. Advances in Appl. Probab. 19:202-218.