REMARKS ON THE EXISTENCE OF SOLUTIONS TO THE AVERAGE COST OPTIMALITY EQUATION IN MARKOV DECISION PROCESSES

Emmanuel Fernández-Gaucherand, Aristotle Arapostathis, and Steven I. Marcus

Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, Texas 78712-1084

Abstract: Necessary conditions are given for the existence of a bounded solution to the optimality equation arising in Markov decision processes, under a long-run, expected average cost criterion. The relationships of some of our results to known sufficient conditions are also shown.

Key Words: Markov decision processes, Borel state and action spaces, average cost optimality equation, bounded solutions, necessary conditions.

1. Introduction

Markov decision processes (MDP) with an infinite planning horizon find important applications in many diverse disciplines. When a discounted cost criterion is used, and the one-stage cost function is bounded, the corresponding dynamic programming operator exhibits nice contractive properties which enable the development of a rather complete theory, under very general conditions [3], [10]. The same cannot be said of the average cost criterion. In this situation, usually a bounded solution to the average cost optimality equation (ACOE) is sought, the existence of which leads to, e.g., optimal stationary policies. However, the controlled process must exhibit a sufficiently stable asymptotic behavior for the above to hold. Thus the type of sufficient conditions available in the literature to guarantee the existence of such bounded solutions imposes strong recurrence restrictions on the model, c.f. [5], [6], [10], [11], [16], [20]. In this note, *necessary* conditions for the existence of a bounded solution to the ACOE are presented. Our results hold under very general

conditions, e.g. Borel state space, compact admissible action sets, and bounded below lower semicontinuous cost function. Furthermore, the average cost is allowed to depend on the initial state. We show that, e.g., if a bounded solution to the ACOE exists, for a given cost function, then necessarily differences of discounted costs are bounded, uniformly in the initial states. Furthermore, if the average cost does not depend on the initial state, the latter holds uniformly with respect to the discount factor. Thus, the necessary conditions presented here complement known *sufficient* conditions, as examined in, e.g., [2], [5], [10], [6], [11], [17], [20]. It is also noted that for some *countable* state space MDP and *finite* (core) state space *partially observable* MDP, some of the conditions presented are both necessary and sufficient for the existence of bounded solutions to the ACOE.

2. Notation and Preliminaries

Given a topological space $\mathbf{W}$, its Borel σ-algebra will be denoted by $\mathcal{B}(\mathbf{W})$. Following similar notation to that in [10], let $\{\mathbf{X}, \mathbf{U}, Q, c\}$ denote a MDP, where the state space $\mathbf{X}$ is a Borel space, i.e. a Borel subset of a complete separable metric space; $\mathbf{U}$ denotes the control or action set, also taken as a Borel space. To each $x \in \mathbf{X}$, a nonempty compact set $\mathbf{U}(x) \in \mathcal{B}(\mathbf{U})$ of admissible actions is associated. Let $\mathbf{K} := \{(x,u) : x \in \mathbf{X}, u \in \mathbf{U}(x)\}$ denote the space of admissible state-action pairs, which is viewed as a topological subspace of $\mathbf{X} \times \mathbf{U}$. The evolution of the system is governed by the *stochastic kernel* Q on $\mathbf{X}$ given $\mathbf{K}$, i.e., $Q(B|\cdot)$ is a Borel measurable function on $\mathbf{K}$, for each $B \in \mathcal{B}(\mathbf{X})$, and $Q(\cdot\,|x,u)$ is a probability measure on $\mathcal{B}(\mathbf{X})$, for each $(x,u) \in \mathbf{K}$. Finally, $c : \mathbf{K} \to I\!R$ is the Borel measurable one-stage cost function. Thus, at time $t \in I\!N_0 := \{0, 1, 2, \ldots\}$, the system is observed to be in some state, say $x \in \mathbf{X}$, and a decision $u \in \mathbf{U}(x)$ is taken. Then a cost $c(x,u)$ is accrued, and by the next decision epoch $t+1$, the state of the system will have evolved to some value in $B \in \mathcal{B}(\mathbf{X})$ with probability $Q(B|x,u)$. The available information for decision-making at time $t \in I\!N_0$ is given by the *history* of the process up to that time $h_t := (x_0, u_0, \ldots, u_{t-1}, x_t) \in \mathbf{H}_t$, where

$$\mathbf{H}_0 := \mathbf{X}, \quad \mathbf{H}_t := \mathbf{H}_{t-1} \times (\mathbf{U} \times \mathbf{X}), \quad \mathbf{H}_\infty := (\mathbf{X} \times \mathbf{U})^\infty$$

are the *history spaces*. With respect to their corresponding product topologies, the above are Borel spaces [3], [10]. An *admissible control policy,* or *strategy,* is

a sequence $\mu = \{\mu_t\}_{t\in\mathbb{N}_0}$ of stochastic kernels μ_t on $\mathbf{U}$ given $\mathbf{H}_t$, satisfying the constraint $\mu_t(\mathbf{U}(x)\,|\,h_t) = 1$, for all $h_t = (h_{t-1}, u, x) \in \mathbf{H}_t$. Of special interest is the set of (nonrandomized) *stationary* policies: if there is a Borel measurable (decision) function $f : \mathbf{X} \to \mathbf{U}$, such that, for all $t \in \mathbb{N}_0$, (i) $f(x) \in \mathbf{U}(x)$, for all $x \in \mathbf{X}$, and (ii) for all $h_t = (h_{t-1}, u, x) \in \mathbf{H}_t$, $\mu_t(\{f(x)\}|h_t) = 1$, then the corresponding policy μ is said to be *stationary*. The set of all stationary policies will be denoted by $\mathcal{S}$. We will simply denote by $\mu(x)$ the action chosen by the stationary policy μ at $x \in \mathbf{X}$. Given the initial state of the process $x \in \mathbf{X}$ and a policy μ, the corresponding state and control processes, $\{X_t\}$ and $\{U_t\}$ respectively, are random processes defined on the canonical probability space $(\mathbf{H}_\infty, \mathcal{B}(\mathbf{H}_\infty), P_x^\mu)$ via the projections $X_t(h_\infty) := x_t$ and $U_t(h_\infty) := u_t$, for each $h_\infty = (x_0, u_0, \ldots, x_t, u_t, \ldots) \in \mathbf{H}_\infty$, where P_x^μ is uniquely determined [3], [10]. The corresponding expectation operator is denoted by E_x^μ.

For a Borel measurable function $v : \mathbf{W} \to \mathbb{R}$, where $\mathbf{W}$ is a topological space, we define

$$\|v\| := \sup_{w\in\mathbf{W}} \{|v(w)|\}, \tag{1}$$

$$\begin{aligned} sp(v) :&= \sup_{w,w'\in\mathbf{W}} \{v(w) - v(w')\} \\ &= \sup_{w\in\mathbf{W}} \{v(w)\} - \inf_{w'\in\mathbf{W}} \{v(w')\}. \end{aligned} \tag{2}$$

Correspondingly, we denote the vector space of bounded, Borel measurable functions $v : \mathbf{W} \to \mathbb{R}$ by

$$\mathcal{M}_b(\mathbf{W}) := \big\{v : \mathbf{W} \to \mathbb{R} \;\big|\; v \text{ is Borel measurable, } \|v\| < \infty\big\}.$$

Hence for $v \in \mathcal{M}_b(\mathbf{W})$, $\|v\|$ and $sp(v)$ give the *supremum norm* and the *span seminorm* of v, respectively. It is easy to check that $sp(v) \le 2\|v\|$, for all $v \in \mathcal{M}_b(\mathbf{W})$. If $v \in \mathcal{M}_b(\mathbf{W})$, then define

$$v^+ := v - \inf_{w\in\mathbf{W}} \{v(w)\} \tag{3}$$

$$v^- := v - \sup_{w\in\mathbf{W}} \{v(w)\}, \tag{4}$$

and note that $sp(v) = v^+(w) - v^-(w)$, for all $w \in \mathbf{W}$. Due to (3) and (4), we conveniently denote the supremum of $v \in \mathcal{M}_b(\mathbf{W})$ as $v - v^-$, and its infimum as $v - v^+$. Also, $\mathcal{L}(\mathbf{W})$ will denote the collection of lower semicontinuous bounded below functions $f : \mathbf{W} \to \mathbb{R}$, and $\mathcal{L}_b(\mathbf{W}) := \mathcal{L}(\mathbf{W}) \cap \mathcal{M}_b(\mathbf{W})$.

3. The Average Cost Optimality Equation

The following two assumptions will be used subsequently, and are in effect throughout, the second of which is made to guarantee the existence of "measurable selectors," c.f. [3, Section 7.5].

Assumption 3.1: There exists $L \in \mathbb{R}$ such that $L \leq c(x,u)$, for all $(x,u) \in \mathbf{K}$.

Assumption 3.2: One of the following holds:
(i) For each $x \in \mathbf{X}$, $\mathbf{U}(x)$ is a finite set; or
(ii) $c(\cdot\,,\cdot) \in \mathcal{L}(\mathbf{K})$, and $\int f(y)Q(dy\,|\,\cdot\,,\cdot) \in \mathcal{L}(\mathbf{K})$, for each $f(\cdot) \in \mathcal{L}(\mathbf{X})$.

Remark 3.1: A sufficient condition for $\int f(y)Q(dy\,|\,\cdot\,,\cdot) \in \mathcal{L}(\mathbf{K})$ to hold, for each $f(\cdot) \in \mathcal{L}(\mathbf{X})$, is that Q is *weakly continuous*, i.e., $\int u(y)Q(dy\,|\,\cdot\,,\,\cdot)$ is a continuous function of $\mathbf{K}$, for all continuous and bounded functions $u : \mathbf{X} \to \mathbb{R}$, c.f. [3, Ch.7].

For an initial state x the *discounted cost* (DC) accrued by policy μ, using a discount factor $0 < \beta < 1$, is given by

$$J_\beta(x,\mu) := \lim_{n\to\infty} E_x^\mu \left[\sum_{t=0}^{n} \beta^t c(X_t, U_t)\right],$$

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

$$J_\beta^*(x) := \inf_\mu \big\{ J_\beta(x,\mu) \big\},$$

the infimum being taken over all admissible policies. Similarly, the long-run expected average cost (AC) accrued by policy μ is given by

$$J(x,\mu) := \limsup_{n\to\infty} \frac{1}{n+1} E_x^\mu \left[\sum_{t=0}^{n} c(X_t, U_t)\right],$$

and the optimal average cost is defined as

$$J^*(x) := \inf_\mu \{ J(x,\mu) \}.$$

If a policy μ is such that $J_\beta(x,\mu) = J_\beta^*(x)$, for all $x \in \mathbf{X}$, then it is said to be DC optimal; AC optimal policies are similarly defined.

Remark 3.2: In view of Assumption 3.1, with no loss in generality, costs may be taken as nonnegative when considering either the DC or AC optimal control problems, as given above.

The *undiscounted* dynamic programming map $T : \mathcal{L}(\mathbf{X}) \to \mathcal{L}(\mathbf{X})$, is defined as

$$T(f)(x) := \inf_{u \in \mathbf{U}(x)} \left\{ c(x,u) + \int_{\mathbf{X}} f(y) Q(dy|x,u) \right\}, \quad \forall x \in \mathbf{X}, \tag{5}$$

and for $0 < \beta < 1$, the *discounted* dynamic programming map T_β is given as

$$T_\beta(f) := T(\beta f). \tag{6}$$

These maps, as well as their iterates, are well defined, c.f., [3, p. 148-149], and the following properties can be immediately verified from (5).

Lemma 3.1: Let $f, f' \in \mathcal{L}(\mathbf{X})$. Then
(i) for all $k \in I\!R$, $T(f + k) = T(f) + k$;
(ii) if $f \leq f'$, then $Tf \leq Tf'$.

Under Assumptions 3.1 and 3.2, the *discounted cost optimality equation* (DCOE) holds:

$$\begin{aligned} J^*_\beta(x) &= \inf_{u \in \mathbf{U}(x)} \left\{ c(x,u) + \beta \int_{\mathbf{X}} J^*_\beta(y) Q(dy|x,u) \right\} \\ &= T_\beta(J^*_\beta)(x), \quad \forall x \in \mathbf{X}. \end{aligned} \tag{7}$$

Furthermore, a stationary policy $\mu \in \mathcal{S}$ is DC optimal if and only if $\mu(x)$ attains the infimum in (7), for all $x \in \mathbf{X}$, and one such policy exists [2], [3], [10]. Note that $J^*_\beta(x) = +\infty$ is not ruled out, and that J^*_β is not necessarily the unique fixed point of T_β, as is the case when $c(\cdot\,,\cdot) \in \mathcal{L}_b(\mathbf{K})$ [3], [10].

If there are Borel measurable real-valued functions Γ and h on $\mathbf{X}$, with $h \in \mathcal{L}(\mathbf{X})$, such that

$$\begin{aligned} \Gamma(x) + h(x) &= \inf_{u \in \mathbf{U}(x)} \left\{ c(x,u) + \int_{\mathbf{X}} h(y) Q(dy|x,u) \right\} \\ &= T(h)(x), \qquad \forall x \in \mathbf{X}, \end{aligned} \tag{8}$$

then the pair (Γ, h) is said to be a *solution of the ACOE*. While the situation involving discounted, possibly unbounded, costs corresponding to (7) is very well understood, e.g. see [2], [3], [10], [14], [17], quite the opposite is true for the average cost case, even for bounded costs [2], [10], [14], [17]. Indeed, the study of MDP with an average cost criterion is an active area of research, e.g. see [4], [5], [6], [7], [8],

[11], [13], [18], [19]. The interest in finding conditions that guarantee solutions to (8) derives from the following result.

Theorem 3.1: Suppose that (Γ, h) is a solution to the ACOE, and that for each admissible policy μ the following holds

$$\lim_{n\to\infty} E_x^{\mu}\left[\frac{h(X_n)}{n}\right] = 0, \qquad \forall x \in \mathbf{X}. \tag{9}$$

Then

(i) $$\limsup_{n\to\infty} \frac{1}{n+1} E_x^{\mu}\left[\sum_{t=0}^{n} \Gamma(X_t)\right] \leq J(x,\mu),$$

and if $\mu \in \mathcal{S}$ is such that $\mu(x)$ attains the infimum in (8), equality is attained above; (ii) if $\Gamma(x) = \Gamma^* \in I\!R$, for all $x \in \mathbf{X}$, then $\Gamma^* = J^*(x)$, for all $x \in \mathbf{X}$, any $\mu^* \in \mathcal{S}$ such that $\mu^*(x)$ attains the infimum in (8) is AC optimal, and one such policy exists.

The proof of Theorem 3.1 is a simple extension of Theorem 2.2 in [9, p.53-55], and will not be given here. Note that (i) above says that if Γ is taken as the cost function to define the MDP $\{\mathbf{X}, \mathbf{U}, Q, \Gamma\}$ then, *for any* admissible policy μ, the average cost assessed under the cost function Γ does not exceed that under cost function c.

Given the results above, naturally there has been considerable interest in finding conditions which guarantee the existence of a *bounded* solution (Γ^*, h) to the ACOE, with $\Gamma^* \in I\!R$ and $h \in \mathcal{L}_b(\mathbf{X})$, for then (9) is satisfied trivially, and (ii) applies. However, the type of sufficient conditions available in the literature, for such a solution to the ACOE to exist, impose a very restrictive recurrence structure on the model under *every* stationary policy, see [5], [6], [10], [11], [16], [20]. For the case of countable state space MDP and bounded costs, Cavazos-Cadena has shown in [6] that the usual conditions used for the above effect are extremely restrictive, in that these not only guarantee the existence of a bounded solution to the ACOE for *any* cost function $c \in \mathcal{L}_b(\mathbf{K})$, but they also do so for a whole *family* of MDP. For general state space MDP, Hernández-Lerma et.al. [11] have given a comprehensive account of recurrence conditions used for the purpose above, and relations among them. Also, given a bounded solution (Γ^*, h) to the ACOE, properties of policies $\mu^* \in \mathcal{S}$ attaining the infimum have been recently investigated in [12]; see also [9]. Our objective is to exhibit some *necessary* conditions that complement known *sufficient* conditions, as examined in, e.g., [2], [5], [6], [7], [8], [10], [11], [12], [14], [16], [17], [20].

4. The Necessary Conditions

Let T_β^k denote the k^{th} iterate of the discounted dynamic programming operator (6). Note that given $v \in \mathcal{L}_b(\mathbf{X})$, then $-\|v\| \leq v \leq \|v\|$, and hence, by Lemma 3.1,

$$T_\beta(0) - \beta\|v\| \leq T_\beta(v) \leq T_\beta(0) + \beta\|v\|.$$

Then, under Assumptions 3.1 and 3.2, the following can be shown [2], [3], [10].

Lemma 4.1: For any $v \in \mathcal{L}_b(\mathbf{X})$, $T_\beta^k(v)(x) \underset{k\to\infty}{\longrightarrow} J_\beta^*(x)$, for all $x \in \mathbf{X}$.

Suppose that (Γ, h) is a solution to the ACOE, and let $k \in I\!R$; then for $\overline{h}(x) := h(x) + k$, we have that $(\Gamma, \overline{h})$ is also a solution to the ACOE, as is easily seen from (8). Thus (Γ, h^+) and (Γ, h^-) are also solutions to the ACOE, if (Γ, h) is a solution with $h \in \mathcal{L}_b(\mathbf{X})$. The following result has been proved by Platzman [15] for *partially observable* MDP , with both $\mathbf{X}$ and $\mathbf{U}$ finite; nevertheless, in this more general setting the proof follows along similar lines.

Lemma 4.2: Suppose that (Γ, h) is a solution to the ACOE, with $\Gamma \in \mathcal{M}_b(\mathbf{X})$ and $h \in \mathcal{L}_b(\mathbf{X})$. Then, for all $x \in \mathbf{X}$,

$$h^-(x) \leq J_\beta^*(x) - \frac{\Gamma - \Gamma^+}{1-\beta}$$

and

$$h^+(x) \geq J_\beta^*(x) - \frac{\Gamma - \Gamma^-}{1-\beta}.$$

Proof: For $0 < \beta < 1$, note that $0 \leq \beta h^+ \leq h^+$, and $h^- \leq \beta h^- \leq 0$. Hence, by (6) and Lemma 3.1,

$$T(h^-) \leq T(\beta h^-) = T_\beta(h^-).$$

Since (Γ, h^-) is also a solution to the ACOE then, for all $x \in \mathbf{X}$, we have

$$\begin{aligned} &\Gamma(x) + h^-(x) \leq T_\beta(h^-)(x) \\ &\quad \Rightarrow h^-(x) \leq T_\beta(h^-)(x) - \frac{1-\beta}{1-\beta}(\Gamma - \Gamma^+), \qquad \forall x \in X. \end{aligned} \tag{10}$$

Proceeding by induction, suppose that for some $k \in I\!N$

$$h^-(x) \leq T_\beta^k(h^-)(x) - \frac{1-\beta^k}{1-\beta}(\Gamma - \Gamma^+). \tag{11}$$

Multiplying both sides of (11) by β, using the first inequality in (10) and Lemma 3.1, we obtain

$$\begin{aligned}\Gamma(x) + h^-(x) &\leq T(\beta h^-)(x) \\ &\leq T(\beta T_\beta^k(h^-))(x) - \frac{\beta - \beta^{k+1}}{1-\beta}(\Gamma - \Gamma^+).\end{aligned}$$

Since $T(\beta T_\beta^k) = T_\beta^{k+1}$, then after rearranging terms in the inequality above, the following is obtained

$$h^-(x) \leq T_\beta^{k+1}(h^-)(x) - \frac{1-\beta^{k+1}}{1-\beta}(\Gamma - \Gamma^+),$$

completing the induction procedure.

Similarly, we obtain that

$$T_\beta^k(h^+)(x) - \frac{1-\beta^k}{1-\beta}(\Gamma - \Gamma^-) \leq h^+(x), \quad \forall k \in I\!N.$$

Hence, taking limits as $k \to \infty$, the result is obtained, by Lemma 4.1. ■

Remark 4.1: If $\Gamma(x) = \Gamma^* \in I\!R$, for all $x \in \mathbf{X}$, then from Lemma 4.2, it is obtained that $(1-\beta)J_\beta^*(x) \to \Gamma^*$, uniformly in x, as $\beta \uparrow 1$.

Our main results can now be easily proved.

Theorem 4.1: Suppose that (Γ, h) is a solution to the ACOE, with $\Gamma \in \mathcal{M}_b(\mathbf{X})$ and $h \in \mathcal{L}_b(\mathbf{X})$. Then
(i) $J_\beta^* \in \mathcal{L}_b(\mathbf{X})$, for all $0 < \beta < 1$, and

$$\left|J_\beta^*(x) - J_\beta^*(y)\right| \leq sp(J_\beta^*) \leq 2sp(h) + \frac{sp(\Gamma)}{1-\beta}, \qquad \forall x, y \in \mathbf{X}, \quad \forall 0 < \beta < 1.$$

Furthermore if (a) for every choice of cost function $c \in \mathcal{L}_b(\mathbf{K})$ there is a corresponding solution to the ACOE (Γ_c, h_c), with $\Gamma_c \in \mathcal{M}_b(\mathbf{X})$ and $h_c \in \mathcal{L}_b(\mathbf{X})$, and (b) there exists $0 < M < \infty$ such that

$$\|h_c\| \leq M\|c\|, \qquad \forall c \in \mathcal{L}_b(\mathbf{K}),$$

then

(ii) $\left|J_{\beta,c}^*(x) - J_{\beta,c}^*(y)\right| \leq 4M\|c\| + \frac{sp(\Gamma_c)}{1-\beta}, \quad \forall x, y \in \mathbf{X}, \forall 0 < \beta < 1, \forall c \in \mathcal{L}_b(\mathbf{X}),$

where $J^*_{\beta,c}$ denotes the DC value function corresponding to the cost function c.

Proof: (i) Under Assumptions 3.1 and 3.2, $J^*_\beta \in \mathcal{L}(\mathbf{X})$ [3]. It then follows immediately from Lemma 4.2 that $J^*_\beta \in \mathcal{L}_b(\mathbf{X})$, since v^+ and v^- are both bounded. Now, let $x, y \in \mathbf{X}$ and $0 < \beta < 1$ be chosen arbitrarily; then

$$\begin{aligned}
\left|J^*_\beta(x) - J^*_\beta(y)\right| &\leq sp(J^*_\beta) \\
&= \sup_{x',y'} \left\{ \left[J^*_\beta(x') - \frac{\Gamma - \Gamma^-}{1-\beta} \right] - \left[J^*_\beta(y') - \frac{\Gamma - \Gamma^+}{1-\beta} \right] \right\} + \frac{sp(\Gamma)}{1-\beta} \\
&\leq \sup_{x',y'} \left\{ h^+(x') - h^-(y') \right\} + \frac{sp(\Gamma)}{1-\beta} \\
&= \sup_{x',y'} \left\{ h(x') - h(y') + sp(h) \right\} + \frac{sp(\Gamma)}{1-\beta} \\
&= 2sp(h) + \frac{sp(\Gamma)}{1-\beta},
\end{aligned}$$

where Lemma 4.2 was used to obtain the second inequality.

(ii) Since

$$sp(h_c) \leq 2\|h_c\| \leq 2M\|c\|,$$

then the result directly follows from (i) above. ■

From (i) in Theorem 4.1, we see that the existence of a bounded solution to the ACOE necessarily imposes the boundedness condition $J^*_\beta \in \mathcal{L}_b(\mathbf{X})$. Usually, a solution with $\Gamma(x) = \Gamma^* \in I\!R$, for all $x \in \mathbf{X}$, is required, giving that $J^*(x) = \Gamma^*$ independently of the initial state. For this case $sp(\Gamma) = 0$, and (i) in Theorem 4.1 implies then that $\{J^*_\beta(x) - J^*_\beta(y)\}$ is *uniformly bounded*, over $x, y \in \mathbf{X}$ and $0 < \beta < 1$. For the case when the state space $\mathbf{X}$ is countable, the action set $\mathbf{U}$ is finite, and a cost function $c \in \mathcal{L}_b(\mathbf{K})$ is used, this uniform boundedness condition is well known to be also a sufficient condition for the existence of a bounded solution to the ACOE [2], [14], [17]. This has been extended by Sennott [18] to the case when the cost function c is bounded below, but not necessarily bounded above, under the additional assumption that $J^*_\beta(x) < \infty$, for all $x \in \mathbf{X}$. These results are shown using the *vanishing discount* method, i.e. by letting $\beta \uparrow 1$ in the DCOE, c.f., [2], [14], [17], [18]. Also, for $c \in \mathcal{L}_b(\mathbf{K})$, $\mathbf{X}$ a countable set and, e.g., $\mathbf{U}$ a finite set, Cavazos-Cadena [5] has shown, under additional assumptions (see Assumption 1.2 in [5]), that the conditions (a) and (b) stated in Theorem 4.1, with Γ_c a constant, are together equivalent to some very strong recurrence conditions for $\{X_t\}$, under *every*

admissible policy (see condition C_3 in [5]); related issues are also treated in [6], [11]. For $\mathbf{X}$ a general Borel space, a finite action set $\mathbf{U}$, and a bounded cost function $c \in \mathcal{L}_b(\mathbf{K})$, Ross [16] has shown that if, in addition to a uniform boundedness condition, the family $\{h_\beta\}_{\beta\in(0,1)}$ is equicontinuous, where for $x_0 \in \mathbf{X}$ arbitrary but fixed $h_\beta(\cdot) := J^*_\beta(\cdot) - J^*_\beta(x_0)$, then a bounded solution to the ACOE exists. For a *partially observable* MDP, with a finite (core) state space $\mathbf{X}$, it is well known that an equivalent MDP can be associated with it, where the latter has as state space the set of probability distributions on $\mathbf{X}$ [1], [2], [14]. It is shown in [15] that a uniform boundedness condition in the equivalent MDP gives rise to a bounded solution to the ACOE. Actually, it is shown in [9] that $\{h_\beta\}$ is an equicontinuous family, and thus the latter result follows as in [16]. We summarize some of our comments above as follows.

Corollary 4.1: Let the MDP $\{\mathbf{X}, \mathbf{U}, Q, c\}$ be such that $\mathbf{U}$ is a finite set, $c \in \mathcal{L}_b(\mathbf{K})$, and either (a) the state space is countable, or (b) it is the equivalent MDP associated with a partially observable MDP with finite (core) state space . Then the following are equivalent:
(i) there exists $M \in I\!R$ such that

$$\left|J^*_\beta(x) - J^*_\beta(y)\right| \leq M, \qquad \forall x, y \in \mathbf{X}, \quad \forall 0 < \beta < 1;$$

(ii) there exists a solution (Γ^*, h) to the corresponding ACOE, with $\Gamma^* \in I\!R$ and $h \in \mathcal{L}_b(\mathbf{X})$.

5. Conclusions

Although it is a classical problem, the MDP with a long-run expected average cost criterion is still far from being completely understood. In this note, we have presented necessary conditions for the existence of a bounded solution to the ACOE. For some situations, some of these conditions are also sufficient for the existence of such a solution, as noted in Corollary 4.1. Thus, our results add a new interesting facet to the understanding of these problems. In particular, it can now be appreciated more clearly how restrictive it is to require bounded solutions to the ACOE, which in turn motivates further studies dealing with unbounded solutions to the ACOE, as in, e.g., [4], [8], [18], [19].

Acknowledgements

This research was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-86-0029, in part by the National Science Foundation under Grant ECS-8617860, in part by the Air Force Office of Scientific Research (AFSC) under Contract F49620-89-C-0044, and in part by the Texas Advanced Technology Program under Grants No. 4327 and 003658-093.

References

[1] K.J. Åström, Optimal control of Markov processes with incomplete state information, *J. Math. Anal. Appl.* **10** (1965) 174-205.

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

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

[4] V.S. Borkar, Control of Markov chains with long-run average cost criterion: the dynamic programming equations, *SIAM J. Control Optim.* **27** (1989) 642-657.

[5] R. Cavazos-Cadena, Necessary and sufficient conditions for a bounded solution to the optimality equation in average reward Markov decision chains, *Systems Control Lett.* **10** (1988) 71-78.

[6] R. Cavazos-Cadena, Necessary conditions for the optimality equation in average reward Markov decision processes, *Appl. Math. Opt.* **19** (1989) 97-112.

[7] E. Fernández-Gaucherand, A. Arapostathis, and S.I. Marcus, On partially observable Markov decision processes with an average cost criterion, *Proc. 28th IEEE Conf. on Decision and Control,* Tampa, FL (1989) 1267-1272.

[8] E. Fernández-Gaucherand, A. Arapostathis, and S.I. Marcus, On the average cost optimality equation and the structure of optimal policies for partially observable Markov decision processes, to appear in *Annals of Operations Research, Special Volume on Markov Decision Processes.*

[9] E. Fernández-Gaucherand, Estimation and control of partially observable Markov decision processes, Ph.D. dissertation, Electrical and Computer Engineering Dept., The University of Texas at Austin.

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

[11] O. Hernández-Lerma, R. Montes-de-Oca, and R. Cavazos-Cadena, Recurrence conditions for Markov decision processes with Borel state space: a survey, preprint, 1990.

[12] O. Hernández-Lerma, J.C. Hennet, and J.B. Lasserre, Average cost Markov decision processes: optimality conditions, LAAS-Report 89307, LAAS-CNRS, Toulouse, France, 1989 (to appear in *J. Math. Anal. Appl.*).

[13] M. Kurano, The existence of a minimum pair of state and policy for Markov decision processes under the hypothesis of Doeblin, *SIAM J. Control Optim.* **27** (1989) 296-307.

[14] P.R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control* (Prentice-Hall, Englewood Cliffs, NJ, 1986).

[15] L.K. Platzman, Optimal infinite-horizon undiscounted control of finite probabilistic systems, *SIAM J. Control Optim.* **18** (1980) 362-380.

[16] S.M. Ross, Arbitrary state Markovian decision processes, *Ann. Math. Stat.* **39** (1968) 2118-2122.

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

[18] Linn I. Sennott, Average cost optimal stationary policies in infinite state Markov decision processes with unbounded costs, *Oper. Res.* **37** (1989) 626-633.

[19] A. Shwartz and A.M. Makowski, On the Poisson equation for Markov chains, Report #EE-646, Faculty of Electrical Engineering, Technion: Israel Institute of Technology, 1987.

[20] L.C. Thomas, Connectedness conditions for denumerable state Markov decision processes, in: R. Hartley, L.C. Thomas and D.J. White, Eds., *Recent Developments in Markov Decision Processes* (Academic Press, London, 1980) 181-204.