The average cost optimality equation: a fixed point approach*

Oscar Vega-Amaya

Departamento de Matemáticas

Universidad de Sonora

México

September 2001

Abstract

We are concerned with the *expected average cost* optimal control problem for discrete-time Markov control processes with Borel spaces and possibly unbounded costs. We show, under a Lyapunov stability condition and a growth condition on the costs, the existence of an stationary optimal policy using a well-known fixed point theorem.

2000 Mathematics Subject Classification: 90C40, 93E20.

Key Words and phrases: discrete-time Makov control processes, expected average cost criterion, Lyapunov conditions, fixed point theorems.

1 Introduction

The *expected average cost* (EAC) optimal control problem is among the most studied optimality criteria for discrete-time Markov control processes (MCPs) and there are several approaches to analyze it, for instance, value and policy iteration algorithms, the vanishing discount factor approach,

*Research supported by Consejo Nacional de Investigación Científica y Tecnológica (CONACYT) Grant 28309E

linear programming, the convex approach, etc. (see, for instance, [1], [2], [6], [7],[8], [20], and their extensive bibliographies).

In recent years, some variants of Lyapunov-like stability conditions have been used in several papers to handle the EAC optimal control problem with unbounded costs for MCPs on Borel spaces ([4], [5], [12], [13]) and, more recently, for semi-Markov control processes ([14], [16], [21]), as well as for zero-sum Markov games ([10], [15], [19]). A key property used in all of these papers, is that the imposed stability conditions yield the so-called *weighted geometric ergodicity* for the Markov chains induced by stationary policies. (The weighted geometric ergodicity is a generalization of the standard uniform geometric ergodicity in Markov chain theory; see, [8, Ch. 7] and [17, Ch. 16] for a detailed discussion of these concepts). This fact makes the main difference with our approach since we use "fixed point arguments" and do not need to use, at least explicitly, the W-geometric ergodicity. Fixed point arguments have been used in several previous paper (see, for instance, [6, Lemma 3.5 and Comments 3.7, pp. 59 and 61], [11], [3]), but under a stronger form of Doeblin condition.

In the present paper, we show the existence of an optimal stationary policy for the EAC control problem with *unbounded costs* for MCPs on *Borel spaces* using a new variant of the Lyapunov condition (Assumption 3.2), and a growth condition on the costs (Assumption 3.1), besides the standard continuity/compactness requirements (Assumption 3.4). To do this, we first show that some operator, which is closely related to the *average cost optimality equation* (ACOE), is a contraction (Theorem 3.5); then, we prove, using very simple arguments, that its fixed point solves the (ACOE) which, in turn, yields the existence of EAC stationary optimal policies (Theorem 3.6).

2 Average cost optimality equation

We consider a standard discrete-time Markov control model $(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, Q, C)$ where the *state space* $\mathbf{X}$ and the *control* or *action space* $\mathbf{A}$ are both Borel spaces; for each $x \in \mathbf{X}, A(x)$ is a Borel subset of $\mathbf{A}$ and it denotes the set of *admissible controls* or *actions* for the state x. We assume that the set $\mathbb{K}{:=}\{(x,a) : x \in \mathbf{X}, a \in A(x)\}$ is a Borel subset of $\mathbf{X} \times \mathbf{A}$. The *transition law* $Q(\cdot|\cdot)$ is a stochastic kernel on $\mathbf{X}$ given $\mathbb{K}$, and the *one-step cost* C is a measurable function on $\mathbb{K}$. We denote by $\mathbb{F}$ the class of all Borel measurable functions $f : \mathbf{X} \to \mathbf{A}$ satisfying the constraint $f(x) \in A(x)$ for each $x \in \mathbf{X}$.

A *control policy* $\pi = \{\pi_t\}$ is a sequence of rules to chose admissible controls, that is, $\pi_t(A(x_t)|h_t) = 1$, for each $t = 0, 1, 2, \cdots$ and each history $h_t = (x_0, a_0, x_1, \cdots, x_{t-1}, a_{t-1}, x_t)$ with $a_k \in A(x_k)$ for $k = 0, 1, \cdots t - 1$. The class of all control policies is denoted by Π. A control policy $\pi = \{\pi_t\}$ is said to be a (deterministic) *stationary* policy if there exists $f \in \mathbb{F}$ such that π_t is concentrated on $f(x_t)$ for all history h_t and $t = 0, 1, 2, \cdots$. In this case, following an standard convention, we denote the policy π by f and identify the class of stationary policies with $\mathbb{F}$.

For notational ease, for a measurable function v on $\mathbb{K}$ and $f \in \mathbb{F}$ we write

$$v_f(x) := v(x, f(x)) \quad x \in \mathbf{X}. \tag{1}$$

In particular, for the cost function C and the transition law Q, we have

$$C_f(x) = C(x, f(x)) \qquad \text{and} \qquad Q_f(\cdot|x) = Q(\cdot|x, f(x)) \quad x \in \mathbf{X}. \tag{2}$$

As is well-known, for each policy $\pi \in \Pi$ and "initial" state $x \in \mathbf{X}$, there exists an stochastic processes $\{(x_t, a_t)\}$ and a probability measure P_x^π—which governs the evolution of the processes—both defined on the sample space $(\Omega, \mathcal{F})$ where $\Omega := (\mathbf{X} \times \mathbf{A})^\infty$ and $\mathcal{F}$ is the product σ-algebra. We will refer to x_t and a_t as the *state* and *control* variables at time t, and denote by E_x^π the expectation operator with respect to the probability measure P_x^π.

When a stationary policy $f \in \mathbf{F}$ is used, the state process $\{x_t\}$ is a Markov chain with one-step probability transition $Q_f(\cdot|\cdot)$. In this case, we denote by $Q^n(\cdot|\cdot), n = 0, 1, \cdots$, the *n-step* probability transition. Thus, in particular, we have for any measurable function u on $\mathbf{X}$ that

$$E_x^f u(x_n) = \int_{\mathbf{X}} u(y) Q_f^n(dy|x) \quad \forall x \in \mathbf{X}, n = 0, 1, \cdots, \tag{3}$$

whenever these expression are well-defined.

The *expected average cost* (EAC) for a policy $\pi \in \Pi$, given the initial state $x_0 = x \in \mathbf{X}$, is defined as

$$J(\pi, x) := \limsup_{n\to\infty} \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} C(x_t, a_t).$$

Thus, the problem we are interested in is to choose a control policy $\pi^* \in \Pi$ with minimum expected average cost, that is,

$$J(\pi^*, x) \leq J(\pi, x) \quad \forall x \in \mathbf{X}, \pi \in \Pi. \tag{4}$$

If a such policy π^* there exists, it is called *expected average cost* (EAC) *optimal.*

Practically all the approaches to solve the EAC optimal control problem are related to find solutions of the average cost optimality equation: a triplet $(\widehat{h}, \widehat{f}, \widehat{\rho})$ where $\widehat{h}$ is measurable function on $\mathbf{X}$, $\widehat{f} \in \mathbb{F}$ and a constant $\widehat{\rho}$, is said to be a solution of the *average cost optimality equation* (ACOE) —or a *canonical triplet*— if for all $x \in \mathbf{X}$ it holds that

$$\widehat{\rho} + \widehat{h}(x) = \min_{a\in A(x)} \left[C(x,a) + \int_{\mathbf{X}} \widehat{h}(y) Q(dy|x,a) \right] = C_{\widehat{f}}(x) + \int_{\mathbf{X}} \widehat{h}(y) Q_{\widehat{f}}(dy|x). \tag{5}$$

The connection between the EAC optimal control problem in (4) and the ACOE in (5) is well-known: if $(\widehat{h}, \widehat{f}, \widehat{\rho})$ solves the ACOE and, in addition, it holds that

$$\lim_{n\to\infty} \frac{1}{n} E_x^\pi |\widehat{h}(x_n)| = 0 \quad \forall \pi \in \Pi, x \in \mathbf{X}, \tag{6}$$

then—by standard dynamic programming arguments— the policy $\widehat{f}$ is EAC optimal and $\widehat{\rho}$ is the EAC optimal cost; that is, $\widehat{\rho} = J(\widehat{f}, x) \leq J(\pi, x)$ for all $\pi \in \Pi$ and $x \in \mathbf{X}$.

We prove in Theorem 3.6, under Assumption 3.1 (growth condition on the costs), Assumption 3.2 (Lyapunov stability condition) and Assumption 3.4 (compactness/continuity requirements), the existence of an EAC optimal stationary policy by showing the existence of a solution of the ACOE satisfying (6).

3 Main results and assumptions

Our first hypothesis imposes a growth condition on the cost function C.

Assumption 3.1. There exists a measurable function $W(\cdot)$ on $\mathbf{X}$ bounded below by a constant $\theta > 0$ and such that $|C(x,a)| \leq KW(x)$ for all $x \in \mathbf{X}$, where K is a positive constant.

To state our second set of hypothesis, we introduce some notation. For a measurable function $u(\cdot)$ on $\mathbf{X}$, we define the *weighted norm* (W–norm, for short) as $||u||_W := \sup_{x\in\mathbf{X}} \frac{|u(x)|}{W(x)}$, and denote by $B_W(\mathbf{X})$ the Banach space of all measurable function with finite W–norm. Moreover, for a measure $\gamma(\cdot)$ on $\mathbf{X}$ let $\gamma(u) := \int_{\mathbf{X}} u(x)\gamma(dx)$ whenever the integral is well-defined.

Assumption 3.2. There exists a non-trivial measure $\nu(\cdot)$ on $\mathbf{X}$, a nonnegative measurable function $\phi(\cdot,\cdot)$ on $\mathbb{K}$ and a positive constant $\lambda < 1$ such that:

(a) $\nu(W) < \infty$;

(b) $Q(B|x,a) \geq \nu(B)\phi(x,a) \quad \forall B \in \mathcal{B}(\mathbf{X}), (x,a) \in \mathbb{K}$;

(c) $\displaystyle\int_{\mathbf{X}} W(y)Q(dy|x,a) \leq \lambda W(x) + \phi(x,a)\nu(W)$;

(d) $\nu(\phi_f) > 0 \quad \forall f \in \mathbb{F}$.

Assumption 3.2 was previously used for Markov and semi-Markov control processes on Borel spaces with unbounded costs in [4], [5] and [16], respectively, but they also impose some additional conditions. Specifically, they also require that the following holds:

Condition I: $\inf_{f\in\mathbb{F}} \nu(\phi_f) > 0$;

Condition II: The whole family of transition laws $Q_f(\cdot|\cdot), f \in \mathbb{F}$, admits a *common* irreducibility measure $\gamma(\cdot)$.

The key consequences of Assumptions 3.1, 3.2 and Conditions I –II can be stated [using the relation (3)] as follows ([4], [16]):

(A) For each $f \in \mathbb{F}$, the transition law $Q_f(\cdot|\cdot)$ is *positive Harris recurrent.* Thus, it admits a *unique invariant probability measure* $\mu_f(\cdot)$, that is, $\mu_f(\cdot) = \int_{\mathbf{X}} Q_f(\cdot|x)\mu_f(dx)$.

(B) There exist positive constants M and $\beta < 1$ such that

$$\sup_{f\in\mathbb{F}} ||Q_f^n u - \mu_f(u)||_W \leq ||u||_W M\beta^n \quad \forall u \in B_W(\mathbf{X}), n = 0, 1, \cdots . \tag{7}$$

(C) Thus, for each $f \in \mathbb{F}$, the function

$$h_f(x) := \sum_{n=0}^{\infty} \left[Q_f^n C_f(x) - \mu_f(C_f)\right] \quad x \in \mathbf{X}, \tag{8}$$

belongs to $B_W(\mathbf{X})$ and it satisfies the *Poisson equation*

$$h_f(x) = C_f(x) - \mu_f(C_f) + \int_{\mathbf{X}} h_f(y) Q_f(dy|x) \quad \forall x \in \mathbf{X}. \tag{9}$$

The approach used in the present paper is quite different since we use "fixed-point arguments", instead of the *W-geometric ergodicity* in (7), to obtain solutions to the Poisson equations and, moreover, we do not need to impose Conditions I-II. In fact, we show that Assumption 3.2 implies that each transition probability $Q_f(\cdot|\cdot), f \in \mathbb{F}$, is ν–irreducible and also that it is positive Harris recurrent. These facts are stated in the following theorem and proved in Section 4.

Theorem 3.3. Under Assumptions 3.1 and 3.2 the following holds. For each $f \in \mathbb{F}$:

(a) $Q_f(\cdot|\cdot)$ is ν–irreducible and positive Harris recurrent with unique invariant probability measure $\mu_f(\cdot)$;

(b) $\mu_f(W) < \infty$; thus

$$\rho_f := \mu_f(C_f) < \infty \quad \text{and} \quad \rho^* := \inf_{f\in\mathbb{F}} \rho_f < \infty; \tag{10}$$

(c) There exists a unique function $h_f^0 \in B_W(\mathbf{X})$ that solves the Poisson equation (9) and which satisfies $\nu(h_f^0) = 0$.

Next, for each $u \in B_W(\mathbf{X})$, define

$$\widehat{T}u(x) := \inf_{a\in A(x)} \left[C(x,a) - \rho^* + \int_{\mathbf{X}} u(y)\widehat{Q}(dy|x,a)\right] \qquad \forall x \in \mathbf{X},$$

where

$$\widehat{Q}(B|x,a) := Q(B|x,a) - \nu(B)\phi(x,a) \quad \forall B \in \mathcal{B}(\mathbf{X}), (x,a) \in \mathbb{K}.$$

It is well-known that, in general, $\widehat{\mathbf{T}}u$ need not to be (Borel) measurable. To avoid this problem it is necessary to impose some continuity/compactness conditions on the model. It is worth to mentioning that the measurability problem can be solved in different settings (see, for instance, [7, Theorem 3.5, p.28]). Here, for simplicity, we use the following one:

Assumption 3.4. For each $x \in \mathbf{X}$:

(a) $A(x)$ is a compact subset of $\mathbf{A}$;

(b) $C(x,\cdot)$ is lower semicontinuous on $A(x)$;

(c) $Q(\cdot|x,\cdot)$ is strongly continuous on $A(x)$, that is, the mapping $a \to \int_{\mathbf{X}} u(y)Q(dy|x,a)$ is continuous for each bounded measurable function u on $\mathbf{X}$;

(d) the mapping $a \to \int_{\mathbf{X}} W(y)Q(dy|x,a)$ is continuous;

(e) $\phi(x,\cdot)$ is continuous on $A(x)$.

Notice that, under Assumption 3.2, $\widehat{Q}(\cdot|\cdot,\cdot)$ is a nonnegative kernel on $\mathbf{X}$ given $\mathbb{K}$, and also that Assumption 3.2(c) can be rewritten equivalently as

$$\int_{\mathbf{X}} W(y)\widehat{Q}(dy|x,a) \leq \lambda W(x) \quad \forall (x,a) \in \mathbb{K}. \tag{11}$$

Thus Assumption 3.2 essentially states that $\widehat{Q}(\cdot|\cdot)$ satisfies a certain contraction property. We shall prove that under Assumptions 3.1 and 3.4 the contraction property is transferred to the operator $\widehat{\mathbf{T}}$.

Theorem 3.5. Suppose that Assumptions 3.1, 3.2 and 3.4 hold. Then:

(a) $\widehat{\mathbf{T}}$ is a contraction operator from $B_W(\mathbf{X})$ into itself with modulus λ; hence, by Banach's Fixed Point Theorem, there exists a unique function $h^* \in B_W(\mathbf{X})$ such that $h^* = \widehat{\mathbf{T}}h^*$;

(b) There exists a policy $f^* \in \mathbb{F}$ such that

$$h^*(x) = C_{f^*}(x) - \rho^* + \int_{\mathbf{X}} h^*(y)\widehat{Q}_{f^*}(dy|x) \quad \forall x \in \mathbf{X}, \tag{12}$$

with ρ^* as in (10).

Now, we are ready to state our main result.

Theorem 3.6. Suppose that Assumptions 3.1, 3.2 and 3.4 hold. Then:

(a) The triplet (h^*, f^*, ρ^*) in Theorem 3.5 satisfies the ACOE (5) and

$$J(f^*, x) = \rho^* \leq J(\pi, x) \quad \forall x \in \mathbf{X}, \pi \in \mathbf{\Pi};$$

(b) The functions $h_f^0, f \in \mathbb{F}$, in Theorem 3.3(c) and h^* satisfy

$$h^*(x) = \inf_{f \in \mathbb{F}^*} h_f^0(x) \quad \forall x \in \mathbf{X},$$

where $\mathbb{F}^* := \{f \in \mathbb{F} : f \text{ is EAC-optimal}\}$.

4 Proofs of Theorems 3.3, 3.5 and 3.6.

To prove Theorems 3.3, 3.5 and 3.6 we need some preliminary results, which are collected in Remark 4.1 and Lemmas 4.2, 4.3, and 4.4.

Remark 4.1. Iterations of the inequality in Assumption 3.2(c) yields

$$\theta \leq E_x^\pi W(x_n) \leq \lambda^n W(x) + \frac{\nu(W)}{(1-\lambda)\nu(\mathbf{X})} \qquad \forall x \in \mathbf{X}, \pi \in \Pi, n = 0, 1, \cdots, \tag{13}$$

where θ is the positive constant in Assumption 3.1. Hence

$$\lim_{n \to \infty} \frac{1}{n} E_x^\pi |u(x_n)| = 0 \quad \forall x \in \mathbf{X}, \pi \in \Pi, u \in B_W(\mathbf{X}). \tag{14}$$

Now, for each $v \in B_W(\mathbf{X})$ and $f \in \mathbb{F}$, define

$$L_f^v u(x) := v(x) + \int_{\mathbf{X}} u(y) \widehat{Q}_f(dy|x) = v(x) + \int_{\mathbf{X}} u(y) Q_f(dy|x) - \nu(u)\phi_f(x) \ \ \forall x \in \mathbf{X}.$$

Lemma 4.2. Suppose that Assumption 3.2 holds. Then, for each $v \in B_W(\mathbf{X})$ and $f \in \mathbb{F}$, L_f^v is a contraction operator on $B_W(\mathbf{X})$ into itself with modulus λ. Hence, by Banach's Fixed Point Theorem, there exists a unique function $h_f^v \in B_W(\mathbf{X})$ such that

$$h_f^v(x) = v(x) + \int_{\mathbf{X}} h_f^v(y) Q_f(dy|x) - \nu(h_f^v)\phi_f(x) \quad \forall x \in \mathbf{X}. \tag{15}$$

Proof. Let $v \in B_W(\mathbf{X})$ and $f \in \mathbb{F}$ be arbitrary but fixed. First note that for all $u, w \in B_W(\mathbf{X})$ we have

$$|L_f^v u(x) - L_f^v w(x)| = \left| \int_{\mathbf{X}} u(y) \widehat{Q}_f(dy|x) - \int_{\mathbf{X}} w(y) \widehat{Q}_f(dy|x) \right| \leq \int_{\mathbf{X}} |u(y) - w(y)| \widehat{Q}_f(dy|x) \; \forall x \in \mathbf{X}.$$

Thus, by (11), we see that

$$|L_f^v u(x) - L_f^v w(x)| \leq ||u - w||_W \int_{\mathbf{X}} W(y) \widehat{Q}_f(dy|x) \leq ||u - w||_W \; \lambda W(x) \quad \forall x \in \mathbf{X};$$

hence

$$\left\| L_f^v u - L_f^v w \right\|_W \leq \lambda \quad \forall u, w \in B_W(\mathbf{X}). \blacksquare$$

Lemma 4.3. Suppose that Assumption 3.2 holds. Then, for each $f \in \mathbb{F}$, there exists a constant $k_f \neq 0$ such that

$$\lim_{n \to \infty} \frac{1}{n} E_x^f \sum_{k=0}^{n-1} \phi_f(x_n) = k_f \quad \forall x \in \mathbf{X}.$$

Proof. Choose an arbitrary $f \in \mathbb{F}$ and take $v \equiv 1$ in Lemma 4.2. Then there exists a (unique) function $h_f^1 \in B_W(\mathbf{X})$ such that

$$h_f^1(x) = 1 + \int_{\mathbf{X}} h_f^1(y) Q_f(dy|x) - \nu(h_f^1)\phi_f(x) \quad \forall \in \mathbf{X}.$$

Iteration of this equation yields

$$h_f^1(x) = n - \nu(h_f^1) E_x^f \sum_{k=0}^{n-1} \phi_f(x_k) + E_x^f h_f^1(x_n) \quad \forall x \in \mathbf{X}, n = 0, 1, \cdots.$$

Hence, by Remark 4.1, multiplying by $1/n$ and letting $n \to \infty$ we have

$$\nu(h_f^1) \lim_{n\to\infty} \frac{1}{n} E_x^f \sum_{k=0}^{n-1} \phi_f(x_k) = 1 \quad \forall x \in \mathbf{X},$$

which proves the desire with $k_f = 1/\nu(h_f^1)$.■

Proof of Theorem 3.3. Pick an arbitrary stationary policy $f \in \mathbb{F}$.

(a) By Lemma 4.2, for each $v \in B_W(\mathbf{X})$ there exists a (unique) function $h_f^v \in B_W(\mathbf{X})$ satisfying

$$h_f^v(x) = v(x) + \int_{\mathbf{X}} h_f^v(y) Q_f(dy|x) - \nu(h_f^v)\phi_f(x) \quad \forall x \in \mathbf{X}.$$

Thus, by iteration again, we have

$$h_f^v(x) = E_x^f \sum_{k=0}^{n-1} v(x_n) - \nu(h_f^v) E_x^f \sum_{k=0}^{n-1} \phi_f(x_k) + E_x^f h_f^v(x_n) \quad \forall x \in \mathbf{X}, n = 1, 2, \cdots.$$

Hence, from Remark 4.1 and Lemma 4.3, we obtain

$$\lim_{n\to\infty} \frac{1}{n} E_x^f \sum_{k=0}^{n-1} v(x_k) = \nu(h_f^v) \lim_{n\to\infty} \frac{1}{n} E_x^f \sum_{k=0}^{n-1} \phi_f(x_k) = \nu(h_f^v) k_f < \infty \quad \forall x \in \mathbf{X}.$$

Then by Theorem 3.2 in [9] the following hold:

(i) The transition probability $Q_f(\cdot|\cdot)$ is positive Harris recurrent; thus, in particular, it is irreducible and it has a unique invariant probability measure $\mu_f(\cdot)$;

(ii) moreover, for any bounded measurable function v on $\mathbf{X}$,

$$\lim_{n\to\infty} \frac{1}{n} E_x^f \sum_{k=0}^{n-1} v(x_k) = \mu_f(v) \quad \forall x \in \mathbf{X}, \tag{16}$$

Finally, from [18, Remark 2.1, p.15], we have that the measure $\nu(\cdot)$ is an irreducibility measure.

(b) Taking $v \equiv W$ and proceeding as in the proof of part (a) , we get a function $h_f^W \in B_W(\mathbf{X})$ satisfying

$$\lim_{n\to\infty} \frac{1}{n} E_x^f \sum_{k=0}^{n-1} W(x_k) = \nu(h_f^W) k_f < \infty \quad \forall x \in \mathbf{X}. \tag{17}$$

Now, consider a sequence $\{w_n\}$ of nonnegative bounded measurable functions converging increasingly to W. Then observe that

$$\frac{1}{N}E_x^f\sum_{k=0}^{N-1} w_n(x_k) \le \frac{1}{N}E_x^f\sum_{k=0}^{N-1} W(x_k) \quad \forall x\in \mathbf{X}, N=1,2,\cdots,$$

which, by (16) and (17), implies $\mu_f(w_n)\le \nu(h_f^W)k_f<\infty$ for all $n=1,2,\cdots$. Thus, letting $n\to\infty$, we see that

$$\mu_f(W)\le \nu(h_f^W)k_f<\infty.$$

Finally, (10) follows from Assumption 3.1, which yields that C_f is in $B_W(\mathbf{X})$.

(c) To prove this part, we introduce the following operator: for each $u\in B_W(\mathbf{X})$, define

$$\widehat{T}_f u(x) := C_f(x)-\rho_f+\int_{\mathbf{X}} u(y)\widehat{Q}_f(dy|x) \quad \forall x\in\mathbf{X},$$

where ρ_f is as in (10). Now, as in the proof of Lemma 4.2, it is easy to prove that $\widehat{T}_f$ is a contraction operator from $B_W(\mathbf{X})$ into itself with modulus λ. Then, there exists a unique function $h_f^0\in B_W(\mathbf{X})$ such that

$$h_f^0(x)=\widehat{T}_f h_f^0(x)=C_f(x)-\rho_f+\int_{\mathbf{X}} h_f^0(y)Q_f(dy|x)-\nu(h_f^0)\phi_f(x) \quad \forall x\in\mathbf{X}.$$

Therefore, integrating both sides with respect to the invariant probability measure $\mu_f(\cdot)$, we obtain

$$\nu(h_f^0)\mu_f(\phi_f)=0. \tag{18}$$

On the other hand, integrating both sides of the inequality

$$\int_X W(y)Q_f(dy|x)\le \lambda W(x)+\nu(W)\phi_f(x) \quad x\in\mathbf{X},$$

with respect to $\mu_f(\cdot)$ we see that $0<(1-\lambda)\theta\le(1-\lambda)\mu_f(W)\le\nu(W)\mu_f(\phi_f)<\infty$, which implies

$$\mu_f(\phi_f)\ge\frac{(1-\lambda)\theta}{\nu(W)}>0, \tag{19}$$

where θ is the positive constant in Assumption 3.1; hence, by (18), $\nu(h_f^0) = 0$. We thus get the Poisson equation

$$h_f^0(x) = C_f(x) - \rho_f + \int_{\mathbf{X}} h_f^0(y) Q_f(dy|x) \quad \forall x \in \mathbf{X}. \blacksquare$$

Before proving Theorem 3.5 we note the following.

Lemma 4.4. Suppose that Assumptions 3.1, 3.2 and 3.4 hold. Then for each $u \in B_W(\mathbf{X})$ there exists $f \in \mathbb{F}$ such that

$$\widehat{\mathbf{T}}u(x) = C_f(x) - \rho^* + \int_{\mathbf{X}} u(y) \widehat{Q}_f(dy|x) \quad \forall x \in \mathbf{X}; \tag{20}$$

hence $\widehat{\mathbf{T}}u$ is measurable and it belongs to $B_W(\mathbf{X})$.

The proof of Lemma 4.4 is omitted since it follows using standard arguments (see, for instance, [4, Lemma 4.2], or [12, Proposition 2.6]).

Proof of Theorem 3.5. Let u be an arbitrary function in $B_W(\mathbf{X})$ and define

$$Lu(x,a) := C(x,a) - \rho^* + \int_{\mathbf{X}} u(y) \widehat{Q}(dy|x,a) \quad \forall (x,a) \in \mathbb{K}.$$

Note, by (11), that for any pair of functions $u, w \in B_W(\mathbf{X})$ we have

$$|Lu(x,a) - Lv(x,a)| \leq ||u - v||_W \int_{\mathbf{X}} W(y) \widehat{Q}(dy|x,a) \leq \lambda ||u - v||_W W(x),$$

for all $(x,a) \in \mathbf{K}$. Thus

$$Lu(x,a) \leq Lv(x,a) + \lambda ||u - v||_W W(x) \quad \forall (x,a) \in \mathbf{K};$$

hence, taking the infimum on $A(x)$ in both sides of this inequality, we see that

$$\widehat{\mathbf{T}}u(x) \leq \widehat{\mathbf{T}}v(x) + \lambda ||u - v||_W W(x) \quad \forall x \in \mathbf{X}.$$

A similar result is obtained, of course, interchanging the role of the functions u and v, and so

$$|\widehat{\mathbf{T}}u(x) - \widehat{\mathbf{T}}v(x)| \leq \lambda ||u - v||_W W(x) \quad \forall x \in \mathbf{X}.$$

Therefore

$$||\widehat{\mathbf{T}}u - \widehat{\mathbf{T}}v||_W \leq \lambda ||u - v||_W.$$

This inequality and Lemma 4.4 show that $\widehat{\mathbf{T}}$ is a contraction operator from $B_W(\mathbf{X})$ into itself with modulus λ.

Part (b) is a direct consequence of Lemma 4.4 and part (a).■

Proof of Theorem 3.6. First note, by (19), that

$$\inf_{f \in \mathbb{F}} \mu_f(\phi_f) \geq \frac{(1-\lambda)\theta}{\nu(W)} > 0. \tag{21}$$

Now, let $h^* \in B_W(\mathbf{X})$ and f^* be as in Theorem 3.5(b). Then

$$h^*(x) = C_{f^*}(x) - \rho^* + \int_{\mathbf{X}} h^*(y) Q_{f^*}(dy|x) - \nu(h^*)\phi_{f^*}(x) \quad \forall x \in \mathbf{X}.$$

Integrating both sides of the latter inequality with respect to the invariant probability measure $\mu_{f^*}(\cdot)$ we obtain $\nu(h^*)\mu_{f^*}(\phi_{f^*}) = \rho_{f^*} - \rho^* \geq 0$. Then, by (10) and (21), we see that $\nu(h^*) \geq 0$. On the other hand, note that

$$h^*(x) \leq C_f(x) - \rho^* + \int_{\mathbf{X}} h^*(y) Q_f(dy|x) - \nu(h^*)\phi_f(x) \quad \forall x \in \mathbf{X}, f \in \mathbb{F};$$

integrating again but now with respect to $\mu_f(\cdot)$, we get $\nu(h^*)\mu_f(\phi_f) \leq \rho_f - \rho^*$ for all policy $f \in \mathbb{F}$, which implies that

$$\nu(h^*) \inf_{f \in \mathbb{F}} \mu_f(\phi_f) \leq 0.$$

Thus, by (21), $\nu(h^*) \leq 0$. It follows that $\nu(h^*) = 0$, and so the triplet (h^*, f^*, ρ^*) satisfies the ACOE.

Finally, by (14), standard dynamic programming arguments yield

$$J(f^*, x) = \rho_{f^*} = \rho^* \leq J(\pi, x) \quad \forall x \in \mathbf{X}, \pi \in \Pi.$$

(b) First note that $h^* = h^0_{f^*}$ since h^* is also a fixed point of the operator $\widehat{T}_{f^*}$ Now, consider a policy $f \in \mathbb{F}^* = \{f \in \mathbb{F} : f \text{ is EAC-optimal}\}$ and let h^0_f be as in Theorem 3.3(c). Then

$$h^0_f(x) = \widehat{T}_f h^0_f(x) \geq \widehat{T} h^0_f(x) \quad \forall x \in \mathbf{X}.$$

This inequality implies that

$$h^0_f(x) \geq \widehat{T}^n h^0_f(x) \quad \forall x \in \mathbf{X}, n = 1, 2, \cdots .$$

Then, since $\widehat{T}^n h^0_f \to h^*$, we have $h^0_f \geq h^* = h^0_{f^*}$. Therefore

$$h^*(x) = \inf_{f \in \mathbb{F}^*} h^0_f(x) \quad \forall x \in \mathbf{X}. \blacksquare$$

Acknowledgment. The author thanks to Professor Onésimo Hernández-Lerma for many useful comments on an early version of this work.

References

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

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

[3] J.I. González-Trejo, O. Hernández-Lerma, L.F. Hoyos-Reyes, *Minimax control of discrete-time stochastic systems*, Reporte Interno 297, Departamento de Matemáticas, CINVESTAV-IPN, México, 2001.

[4] E. Gordienko and O. Hernández-Lerma, *Average cost Markov control processes with weighted norms: existence of canonical policies*, Appl. Math. (Warsaw) 23 (1995) 199-218.

[5] E. Gordienko and O. Hernández-Lerma, *Average cost Markov control processes with weighted norms: value iteration*, Appl. Math. (Warsaw) 23 (1995) 219-237.

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

[7] O. Hernández-Lerma and J.B. Lasserre, *Discrete-Time Markov Control Process: Basic Optimality Criteria*, Springer-Verlag, New York, 1996.

[8] O. Hernández-Lerma and J.B. Lasserre, *Further Topics on Discrete-Time Markov Control Processes*, Springer-Verlag, New York, 1999.

[9] O. Hernández-Lerma and J.B. Lasserre, *Further criteria for positive Harris recurrence of Markov chains*, Proceedings of the American Mathematical Society 129 (2000) 1521-1524.

[10] O. Hernández-Lerma and J.B. Lasserre, *Zero-sum stochastic games in Borel spaces: average payoff criteria*, SIAM J. Control Optim. 39 (2001) 1520–1539

[11] O. Hernández-Lerma, R. Montes-de-Oca, R. Cavazos-Cadena, *Recurrence conditions for MDPs with Borel state space: A survey*, Ann. Oper. Res. 28 (1991) 29-46.

[12] O. Hernández-Lerma and O. Vega-Amaya, *Infinite-horizon Markov control processes with undiscounted cost criteria: from average to overtaking optimality*, Appl. Math. (Warsaw) 25 (1998) 153-178.

[13] O. Hernández-Lerma, O. Vega-Amaya and G. Carrasco, *Sample-path optimality and variance-minimization of average cost Markov control processes*, SIAM J. Control Optim. 38 (1999) 79-93.

[14] A. Jáskiewicz, *An approximation approach to ergodic semi-Markov control processes*, to appear in Math. Methods Oper. Res. 54 (2001).

[15] A. Jáskiewicz and A. Nowak, *On the optimality equation for zero-sum ergodic stochastic games*, to appear in Math. Methods Oper. Res. 54 (2001).

[16] F. Luque-Vásquez and O. Hernández-Lerma, *Semi-Markov control models with average costs*, Appl. Math.(Warsaw) 26 (1999) 315-331.

[17] S.P. Meyn and R.L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, London, 1993.

[18] E. Nummelin, *General Irreducible Markov Chains and Non-negative Operators,* Cambridge University Press, Cambibidge, 1983.

[19] A. Nowak, *Optimal strategies in a class of zero-sum ergodic stochastic games*, Math. Methods Oper. Res. 50 (1999) 399-419.

[20] M.L. Puterman, *Markov Decision Processes: Discrete Sotchastic Dynamic Programming*, Wiley, New York, 1994.

[21] O. Vega-Amaya and F. Luque-Vásquez, *Sample-path average cost optimality for semi-markov control processes on Borel spaces: Unbounded costs and mean holding times*, Appl. Math. (Warsaw) 27 (2000) 343-367.