APPLICATIONES MATHEMATICAE
26,4 (1999), pp. 363–381

O. VEGA-AMAYA (Sonora)

SAMPLE PATH AVERAGE OPTIMALITY OF MARKOV CONTROL PROCESSES WITH STRICTLY UNBOUNDED COST

Abstract. We study the existence of *sample path average cost* (SPAC-) optimal policies for Markov control processes on *Borel* spaces with *strictly unbounded* costs, i.e., costs that grow without bound on the complement of compact subsets. Assuming only that the cost function is lower semicontinuous and that the transition law is weakly continuous, we show the existence of a relaxed policy with "minimal" expected average cost and that the optimal average cost is the limit of discounted programs. Moreover, we show that if such a policy induces a positive Harris recurrent Markov chain, then it is also sample path average (SPAC-) optimal. We apply our results to inventory systems and, in a particular case, we compute explicitly a deterministic stationary SPAC-optimal policy.

1. Introduction. We study the existence of *sample path average cost* (SPAC-) optimal policies for Markov control processes on *Borel* spaces with *strictly unbounded* costs, i.e., costs that grow without bound on the complement of compact subsets. There is a huge literature dealing with the *expected average cost* (EAC) criterion [see Arapostathis *et al.* (1993), Hernández-Lerma and Lasserre (1996) and the references therein], but in contrast, the sample path (or pathwise) analysis is seldom carried out and, when it is done, it is restricted either to the *denumerable* state space case [Borkar (1991), Cavazos-Cadena and Fernández-Gaucherand (1995), Mandl and Lausmanová (1991)] or to *bounded* one-step costs [Arapostathis *et al.* (1993)], in any of these cases, under strong recurrence/ergodicity conditions. To the best of our knowledge, the only works dealing with sample path optimality

1991 *Mathematics Subject Classification*: 93E20, 90C40.

Key words and phrases: Markov control processes, sample path average cost criterion, strictly unbounded costs, inventory systems.

Research partially supported by CONACYT grant 3115P-E9608.

on Borel spaces and unbounded cost are the papers by Hernández-Lerma *et al.* (1998) and Lasserre (1996). It is important to note that the approaches in these papers differ from ours; in fact, roughly speaking, in the former a "V-uniform ergodicity" assumption is used, whereas in the latter the control problem is studied via (infinite-dimensional) linear programming.

In the present paper, assuming solely lower semicontinuity of the one-step cost function and weak continuity of the transition law, we show that the expected and sample path average control problems with strictly unbounded costs are "well-behaved" in the sense that to prove, for every policy and initial distribution, that the SPAC is bounded below by the minimum EAC as well as to ensure the existence of a "relaxed" policy with "minimal" EAC [see Theorems 3.4 and 3.6(a), respectively], it suffices to assume that the EAC is *finite* for some policy and initial distribution. Moreover, we show that if the relaxed policy with minimal cost induces a positive Harris recurrent Markov chain, then it is also SPAC-optimal [Theorem 3.6(b)].

The remainder of the paper is organized as follows. Section 2 contains a brief description of the relevant Markov control model and the assumptions. In Section 3 we introduce the optimality criteria and state the main results [Theorems 3.4–3.6]; their proofs are given in Sections 5 and 6. In Section 4 we discuss several examples from inventory theory and, in a specific case, we compute *explicitly* a (deterministic) stationary policy which is both (strong) EAC-optimal and SPAC-optimal [see Definition 3.2 below].

We shall use the following notation. Given a *Borel space* Y (i.e., a Borel subset of some separable complete metric space), $\mathcal{B}(Y)$ denotes its Borel σ-algebra, and "measurable" will mean "Borel-measurable". $\mathbf{P}(Y)$ stands for the class of all probability measures on Y. Moreover if Y and Z are Borel spaces then a *stochastic kernel* on Y given Z is a function $P(\cdot\,|\,\cdot)$ such that $P(\cdot\,|\,z)$ is a probability measure on Y for each $z \in Z$, and $P(B\,|\,\cdot)$ is a measurable function for each $B \in \mathcal{B}(Y)$. The family of all stochastic kernels on Y given Z is denoted by $\mathbf{P}(Y\,|\,Z)$. Finally, we denote by $\mathbb{N}$ (resp., $\mathbb{N}_0$) the set of positive integers (resp., nonnegative integers).

2. The Markov model. Since the *Markov control model* $(\mathbf{X}, \mathbf{A}, \{A(x) : x \in A(x)\}, Q, C)$ we are concerned with is quite standard, we only give a brief description. For details see, for instance, Hernández-Lerma and Lasserre (1996).

We assume that the state space $\mathbf{X}$ and the control space $\mathbf{A}$ are both Borel spaces. For each $x \in \mathbf{X}$, $A(x)$ is a nonempty Borel subset of $\mathbf{A}$ and, moreover, $\mathbf{K} := \{(x,a) : a \in A(x),\ x \in \mathbf{X}\}$ is a Borel subset of the Cartesian product $\mathbf{X} \times \mathbf{A}$. Finally, the transition law Q is a stochastic kernel on $\mathbf{X}$ given $\mathbf{K}$ and the one-step cost function C is a measurable function on $\mathbf{K}$.

Define

$$\mathbf{H}_0 := \mathbf{X} \quad \text{and} \quad \mathbf{H}_t := \mathbf{K}^t \times \mathbf{X} \quad \text{for } t \in \mathbb{N}.$$

An (admissible) *control policy* is a sequence $\delta = \{\delta_t\}$ such that, for each $t \in \mathbb{N}_0$, $\delta_t \in \mathbf{P}(\mathbf{A} \,|\, \mathbf{H}_t)$ and it satisfies the constraint $\delta_t(A(x_t) \,|\, h_t) = 1$ for all $h_t = (x_0, a_0, \ldots, x_{t-1}, a_t, x_t) \in \mathbf{H}_t$. A control $\delta = \{\delta_t\}$ is said to be: (i) *relaxed* (or randomized stationary) if there exists $\varphi \in \mathbf{P}(\mathbf{A} \,|\, \mathbf{X})$ such that, for each t, $\delta_t(\cdot \,|\, h_t) = \varphi(\cdot \,|\, x_t)$ for all $h_t \in \mathbf{H}_t$; (ii) (deterministic) *stationary* if there exists a measurable function $f : \mathbf{X} \to \mathbf{A}$ such that $f(x) \in A(x)$ for all $x \in \mathbf{X}$, and $\delta_t(\cdot \,|\, h_t)$ is concentrated at $f(x_t)$ for all $h_t \in \mathbf{H}_t$ and $t \in \mathbb{N}_0$.

The class of all control policies is denoted by Δ, while Φ and $\mathbf{F}$ stand for the subclasses formed by the relaxed and stationary policies, respectively.

For each policy $\delta \in \Delta$ and *initial distribution* $\nu \in \mathbf{P}(\mathbf{X})$, there exist a stochastic process $\{(x_t, a_t) : t = 0, 1, \ldots\}$ and a probability measure P_ν^δ—which governs the evolution of the process—both defined on the sample space $(\Omega, \mathcal{F})$, where $\Omega := (\mathbf{X} \times \mathbf{A})^\infty$ and $\mathcal{F}$ is the corresponding product σ-algebra. The expectation operator with respect to P_ν^δ is denoted by E_ν^δ. We will refer to x_t and a_t as the *state* and *control* at time t, respectively. If the initial probability measure ν is concentrated at an initial state $x_0 = x \in \mathbf{X}$, we write P_x^δ and E_x^δ instead of P_ν^δ and E_ν^δ, respectively.

When using a relaxed policy $\varphi \in \Phi$, the state process $\{x_t\}$ is a Markov chain on $\mathbf{X}$ with time-homogeneous transition kernel

$$Q(\cdot \,|\, x, \varphi) := \int_{\mathbf{X}} Q(\cdot \,|\, x, a)\, \varphi(da \,|\, x), \qquad x \in \mathbf{X}. \tag{1}$$

We also write

$$C(x, \varphi) := \int_{\mathbf{X}} C(x, a)\, \varphi(da \,|\, x). \tag{2}$$

For a deterministic stationary policy $f \in \mathbf{F}$, (1)–(2) become

$$Q(\cdot \,|\, x, f) := Q(\cdot \,|\, x, f(x)) \quad \text{and} \quad C(x, f) := C(x, f(x)). \tag{3}$$

We also suppose that the Markov control model has the following properties:

Assumption 2.1. (a) C is nonnegative and lower semicontinuous on $\mathbf{K}$.

(b) C is *strictly unbounded*, i.e., there exists an increasing sequence of compact sets $\mathbf{K}_n \uparrow \mathbf{K}$ such that

$$\lim_{n\to\infty} \inf\{C(x, a) : (x, a) \notin \mathbf{K}_n\} = \infty.$$

(c) $Q(\cdot \,|\, x, a)$ is weakly continuous in $(x, a) \in \mathbf{K}$, i.e., $\int_{\mathbf{X}} u(y)\, Q(dy \,|\, x, a)$ is continuous in $(x, a) \in \mathbf{K}$ for every bounded continuous function u on $\mathbf{X}$.

The property in Assumption 2.1(b) is also referred to by saying that C is a *moment* or that C is a *norm-like function* on $\mathbf{K}$. This assumption has nice consequences, which have been exploited in several contexts [see, for instance, Hernández-Lerma (1993), Hernández-Lerma and Lasserre (1995, 1997), Meyn (1989, 1995), and references therein]. In fact, in Hernández-Lerma (1993), it is shown that Assumptions 2.1 and 3.1 (below) guarantee the existence of a "relaxed" policy which is a "minimum pair" [see Definition 3.2(c) and Theorem 3.6(a) below]. We show this fact again, but our proof exhibits another nice property of the EAC control problem with strictly unbounded costs, namely, that the optimal average cost is the limit of discounted programs [Theorem 3.6(a)]. Moreover, in Theorem 3.6(b), we prove that if such a relaxed policy induces a positive Harris recurrent Markov chain, then it is also SPAC-optimal [see Definition 3.2(d)].

3. Sample path and expected average cost. Our main interest is to evaluate the stochastic control system when a policy $\delta \in \varDelta$ is used, given an initial distribution $\nu \in \mathbf{P}(\mathbf{X})$, by means of the *sample path average cost* (SPAC) defined as

$$J_0(\delta, \nu) := \limsup_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t, a_t), \tag{4}$$

but we also consider the *expected average cost* (EAC) given by

$$J(\delta, \nu) := \limsup_{n\to\infty} \frac{1}{n} E_\nu^\delta \sum_{t=0}^{n-1} C(x_t, a_t). \tag{5}$$

Moreover, we define the *optimal* (minimum) *average cost* as

$$j^* := \inf_\nu \inf_\delta J(\delta, \nu). \tag{6}$$

To avoid a trivial problem we shall use the following assumption.

Assumption 3.1. There exists a policy δ_* and an initial distribution ν_* such that $J(\delta_*, \nu_*)$ is finite.

The optimality criteria we are concerned with are the following.

Definition 3.2. Let δ^* be a policy and ν^* an initial distribution.

(a) δ^* is said to be *expected average cost* (EAC-) *optimal* if

$$J(\delta, x) \geq J(\delta^*, x) \quad \forall x \in \mathbf{X},\ \delta \in \varDelta.$$

(b) δ^* is said to be *strong expected average cost* (strong EAC-) *optimal* if

$$\liminf_{n\to\infty} \frac{1}{n} E_x^\delta \sum_{t=0}^{n-1} C(x_t, a_t) \geq J(\delta^*, x) \quad \forall x \in \mathbf{X},\ \delta \in \varDelta.$$

(c) (δ^*, ν^*) is said to be a *minimum pair* if $J(\delta^*, \nu^*) = j^*$.

(d) δ^* is said to be *sample path average cost* (SPAC-) *optimal* if for every $\delta \in \varDelta$ and $\nu \in \mathbf{P}(\mathbf{X})$,

$$J_0(\delta^*, \nu) = j^* \quad P_\nu^{\delta^*}\text{-almost surely,} \tag{7}$$

and, moreover,

$$J_0(\delta, \nu) \geq j^* \quad P_\nu^{\delta}\text{-almost surely.} \tag{8}$$

Next, we introduce several special classes of policies.

Definition 3.3. A relaxed policy $\varphi \in \varPhi$ is said to be:

(a) *stable* if there exists an *invariant* probability measure $\mu_\varphi \in \mathbf{P}(\mathbf{X})$ for the transition law $Q(\cdot \mid x, \varphi)$, i.e.,

$$\mu_\varphi(\cdot) = \int_{\mathbf{X}} Q(\cdot \mid y, \varphi)\, \mu_\varphi(dy),$$

which satisfies

$$J(\varphi, \mu_\varphi) = \int_{\mathbf{X}} C(y, \varphi)\, \mu_\varphi(dy) < \infty;$$

(b) *Harris recurrent* if there exists a nontrivial σ-finite measure λ_φ on $\mathbf{X}$ such that $\lambda_\varphi(B) > 0$ implies

$$P_x^\varphi[x_t \in B, \text{ for some } t] = 1 \qquad \forall x \in \mathbf{X};$$

(c) *positive Harris recurrent* if it is stable and Harris recurrent.

We denote by $\varPhi_{\mathrm{S}}$ the class of (relaxed) stable policies and by $\varPhi_{\mathrm{R}}$ the class of relaxed policies which are Harris recurrent, while $\varPhi_{\mathrm{P}}$ stands for the class of positive Harris recurrent polices. Note that $\varPhi_{\mathrm{P}} = \varPhi_{\mathrm{S}} \cap \varPhi_{\mathrm{R}}$.

We suppose throughout the following that Assumptions 2.1 *and* 3.1 *hold.*

We now state one of our main results. The proof is given in Section 5.

Theorem 3.4. *For each policy $\delta \in \varDelta$ and measure $\nu \in \mathbf{P}(\mathbf{X})$,*

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t, a_t) \geq j^* \quad P_\nu^\delta\text{-almost surely.} \tag{9}$$

In the next theorem, we obtain as direct consequences of Theorem 3.4 some interesting relations between the concept of minimum pair and the sample path and expected average costs. Part (c) of this theorem was already proved in Hernández-Lerma (1993), but its proof is included here for completeness.

Theorem 3.5. (a) *A policy $\delta^* \in \varDelta$ is EAC-optimal if and only if it is strong EAC-optimal.*

(b) *If* (δ,ν) *is a minimum pair*, *with* $\delta\in\varDelta$ *and* $\nu\in\mathbf{P}(\mathbf{X})$, *then*

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{t=0}^{n-1}C(x_t,a_t)=j^*\qquad P_\nu^\delta\text{-}almost\ surely.$$

(c) *Let* $\varphi\in\varPhi_{\rm S}$ *and* μ_φ *an associated invariant probability measure. Then* (φ,μ_φ) *is a minimum pair if and only if* $J(\varphi,x)=j^*$ *for* (μ_φ-) *almost all* $x\in\mathbf{X}$.

The first part of the next theorem state the existence of a minimum pair (φ^*,μ^*) with φ^* being a stable policy and μ^* an associated invariant probability measure. This result was already proved in Hernández-Lerma (1993), but his approach differs from ours in that his analysis relies on the well-behavior of the expected average cost whereas our analysis is based on the discounted cost. Roughly speaking, our proof of the existence of a minimum pair yields, at the same time, that the optimal average cost may be approximated by discounted programs, which exhibits another nice property of the control problem with strictly unbounded cost. In the second part of the theorem, we show that if the policy φ^* is positive Harris recurrent then it is SPAC-optimal. To state precisely these facts, we introduce the following notation.

For each $\alpha\in(0,1)$, the (expected) *α-discounted cost* under a policy $\delta\in\varDelta$, given the initial distribution measure $\nu\in\mathbf{P}(\mathbf{X})$, is defined by

$$V_\alpha(\delta,\nu):=E_\nu^\delta\sum_{t=0}^{\infty}\alpha^tC(x_t,a_t),\tag{10}$$

and the *α-discounted optimal value* is given by

$$m_\alpha:=\inf_\nu\inf_\delta V_\alpha(\delta,\nu).\tag{11}$$

Theorem 3.6. (a) *There exists a stable policy* $\varphi^*\in\varPhi_{\rm S}$ [*with invariant probability measure* μ^*] *such that* (φ^*,μ^*) *is a minimum pair*; *hence*, *from Theorem* 3.5(c),

$$J(\varphi^*,x)=j^*\qquad for\ \mu^*\text{-}almost\ all\ x\in\mathbf{X}.\tag{12}$$

Moreover,

$$j^*=\lim_{\alpha\to1^-}(1-\alpha)m_\alpha.\tag{13}$$

(b) *If the policy* φ^* *is positive Harris recurrent*, *then it is SPAC-optimal.*

4. Examples. In this section we discuss some examples from inventory theory to illustrate the potential of the approach used in this paper; in fact, in Example B we compute explicitly a (deterministic) stable stationary policy which is both strong EAC- and SPAC-optimal. In Hernández-Lerma

(1993), Hernández-Lerma and Lasserre (1997) and Meyn (1995) other interesting examples are given, including the LQ control problem, which satisfy the assumptions in Theorems 3.4–3.6.

We consider an inventory system with a single product and infinite storage and production capacities, for which the excess demand is not backlogged. Denote by x_t and a_t the inventory level and the amount of product ordered (and immediately supplied) at the beginning of each decision period $t = 0, 1, \ldots$, respectively. The product demand during period t is denoted by w_t, which is assumed to be a nonnegative random variable. The inventory level evolves in $\mathbf{X} = [0, \infty)$ according to

$$(14) \qquad x_{t+1} = (x_t + a_t - w_t)^+, \quad t = 1, 2, \ldots; \ x_0 = x \in \mathbf{X},$$

where $(y)^+ := \max(y, 0)$, and we assume that the production variables $\{a_t\}$ take values in $\mathbf{A} = [0, \infty)$ irrespective of the stock levels, that is, $\mathbf{A} = A(x) := [0, \infty)$ for all $x \in \mathbf{X}$. Moreover, throughout this section we also suppose that the following holds.

ASSUMPTION 4.1. (a) The demand process $\{w_t\}$ is formed by i.i.d. random variables. The common cumulative distribution is denoted by $G(\cdot)$.

(b) $G(y) < 1$ for all $y \geq 0$.

REMARK 4.2. Note that Assumption 4.1(a) implies Assumption 2.1(c), while Assumption 4.1(b) guarantees that any relaxed stable policy is irreducible and Harris recurrent (hence, positive Harris recurrent) with respect to the measure $\lambda(B) := \mathbf{I}_B(0)$, $B \in \mathcal{B}(\mathbf{X})$, where $\mathbf{I}_B(\cdot)$ denotes the indicator function.

In what follows, E denotes the expectation with respect to the joint distribution of the random variables $w_0, w_1, \ldots$

EXAMPLE A. The one-step cost function has the form

$$(15) \qquad C(x, a) = F_1(x + a) + F_2(a),$$

where $F_1(\cdot)$ and $F_2(\cdot)$ are functions from $[0, \infty)$ into itself satisfying:

ASSUMPTION 4.3. (a) $F_1(\cdot)$ and $F_2(\cdot)$ are lower semicontinuous functions bounded from below.

(b) There exist increasing unbounded sequences $\{y_n^1\}$ and $\{y_n^2\}$ of positive numbers such that

$$\lim_{n\to\infty} \inf_{y > y_n^i} F_i(y) = \infty \qquad \text{for } i = 1, 2.$$

(c) $EF_2(\min(y, w_0)) < \infty$ for all $y \geq 0$.

Note that Assumption 4.3 is general enough to include problems with a set-up cost, that is, a fixed cost for placing orders [Bertsekas (1987), Lee and Nahmias (1993)].

REMARK 4.4. (a) A policy $f_K \in \mathbf{F}$ is said to be a *K-threshold* policy if $f_K(x) = K - x$ for $0 \le x \le K$ and $f_K(x) = 0$ for $x > K$. For this policy, direct computations yield

$$J(f_K, x) = F_1(K) + EF_2(\min(K, w_0)) < \infty \quad \forall x \in \mathbf{X}; \tag{16}$$

thus, Assumption 3.1 holds.

(b) Note that, under Assumption 4.1,

$$\mu_K(B) := \int_B (K - w)^+ \, G(dw), \qquad B \in \mathcal{B}(\mathbf{X}),$$

is the unique invariant probability measure for the policy f_K.

THEOREM 4.5. *If Assumptions* 4.1 *and* 4.3 *hold*, *then there exists a relaxed policy* $\varphi^* \in \Phi_P$ *which is SPAC-optimal and*, *moreover*, $J(\varphi^*, x) = j^*$ *for* μ^**-almost all* $x \in \mathbf{X}$.

Proof. It is easy to check that Assumptions 4.1(a) and 4.3 imply Assumption 2.1. Thus, from Remarks 4.2, 4.4 and Theorems 3.5 and 3.6, we see that the assertions in Theorem 4.5 hold. ■

EXAMPLE B. We now consider a particular case of (15) in which we are able to compute *explicitly* a (deterministic) stationary stable policy which is strong expected and sample path average optimal. We take $F_2(y) = by$, $y \ge 0$, where b is a nonnegative constant, so that (15) becomes

$$C(x, a) = F_1(x + a) + ba \qquad \forall (x, a) \in \mathbf{K}. \tag{17}$$

Instead of Assumption 4.3, we now assume that the following hypothesis holds.

ASSUMPTION 4.6. (a) $F_1(\cdot)$ is a convex function bounded from below.

(b) $\lim_{y\to\infty} F_1(y) = \infty$.

Note that, for the specific function $F_2(\cdot)$ we are considering here, Assumption 4.6 implies Assumption 4.3. Hence, under Assumptions 4.1 and 4.6, the results in Theorem 4.5 hold. Next we show that a threshold-type policy is strong expected and sample path average cost optimal. To do this, we define

$$L(y) := F_1(y) + bE\min(y, w_0) \quad \text{for } y \ge 0 \quad \text{and} \quad \varrho^* := \inf_{y\ge 0} L(y). \tag{18}$$

REMARK 4.7. (a) Simple computations yield that for each $K \ge 0$, the K-threshold policy satisfies

$$L(K) = J(f_K, x) \qquad \forall x \in \mathbf{X}.$$

(b) Moreover, there exists $K^* \ge 0$ such that $L(K^*) = \varrho^* = \inf_{y\ge 0} L(y)$; indeed, this follows from the continuity of $L(\cdot)$ and the fact that $\lim_{y\to\infty} L(y) = \infty$.

THEOREM 4.8. *Suppose that Assumptions* 4.1 *and* 4.6 *hold. Then the* K^**-threshold policy is strong expected and sample path average cost optimal, where* K^* *is as in Remark* 4.7(b).

Proof. We require some results on discounted-cost control problems. For each $\alpha \in (0,1)$, recall from (10) that

$$V_\alpha(\delta, x) = E_x^\delta \sum_{t=0}^{\infty} \alpha^t C(x_t, a_t), \qquad x \in \mathbf{X},\ \delta \in \Delta,$$

and define

$$V_\alpha(x) := \inf_{\delta \in \Delta} V_\alpha(\delta, x), \qquad x \in \mathbf{X}. \tag{19}$$

Now, from (12), there exists a stable policy φ^* with invariant probability measure μ^* such that

$$J(\varphi^*, \mu^*) = j^* \quad \text{ for } \mu^*\text{-almost all } x \in \mathbf{X};$$

thus, from a well-known Abelian Theorem [see Hernández-Lerma and Lasserre (1996), Lemma 5.3.1, p. 84],

$$j^* = \lim_{\alpha \to 1^-} (1-\alpha) V_\alpha(\varphi, x) \geq \limsup_{\alpha \to 1^-} (1-\alpha) V_\alpha(x) \quad \text{ for } \mu^*\text{-almost all } x \in \mathbf{X}.$$

Then, since $V_\alpha(\cdot) \geq m_\alpha$ for all $\alpha \in (0,1)$, we see from this and (13) that

$$j^* = \lim_{\alpha \to 1^-} (1-\alpha) V_\alpha(x) \quad \text{ for } \mu^*\text{-almost all } x \in \mathbf{X}. \tag{20}$$

Then, to conclude that the K^*-threshold policy is strong EAC- and SPAC-optimal, it suffices to prove that

$$\varrho^* = \lim_{\alpha \to 1^-} (1-\alpha) V_\alpha(0). \tag{21}$$

In order to do this, first note that

$$V_\alpha(x) \leq V_\alpha(f_K, x), \qquad 0 \leq x \leq K,$$

where f_K is the K-threshold policy; then, taking K large enough we see that $V_\alpha(\cdot) < \infty$ for all $\alpha \in (0,1)$. Now, using Assumption 4.6, it is easy to prove that $V_\alpha(\cdot)$ is a convex function; thus, the function

$$T_\alpha(y) := F_1(y) + by + \alpha E V_\alpha[(y - w_0)^+], \qquad y \geq 0,$$

is convex and $\lim_{y\to\infty} T(y) = \infty$, which implies that there exists a constant $K_\alpha \geq 0$ such that $T_\alpha(K_\alpha) = \inf_{y \geq 0} T_\alpha(y)$. Hence, for each $\alpha \in (0,1)$, $V_\alpha(\cdot)$ satisfies the *α-Discounted Cost Optimality Equation* [Hernández-Lerma and Muñoz-de-Osak (1992)]

$$V_\alpha(x) = \min_{a \in \mathbf{A}} [F_1(x+a) + ba + \alpha E V_\alpha[(x + a - w_0)^+]] \quad \forall x \in \mathbf{X}, \tag{22}$$

and the K_α-threshold policy attains the minimum on the right-hand side of (22), that is, for all $x \in \mathbf{X}$,

$$V_\alpha(x) = F_1(x + f_\alpha(x)) + bf_\alpha(x) + \alpha E V_\alpha[(x + f_\alpha(x) - w_0)^+], \tag{23}$$

where, for each $\alpha \in (0,1)$, f_α denotes the K_α-threshold policy.

Then standard arguments yield

$$V_\alpha(x) = V_\alpha(f_\alpha, x) \quad \forall x \in \mathbf{X},\ \alpha \in (0,1). \tag{24}$$

Moreover, simple computations show that for all $\alpha \in (0,1)$,

$$(1-\alpha)V_\alpha(f_\alpha, 0) = F_1(K_\alpha) + \alpha E \min(K_\alpha, w_0) + b(1-\alpha)K_\alpha. \tag{25}$$

Now define

$$L_\alpha(y) := F_1(y) + \alpha E \min(y, w_0) + b(1-\alpha)y, \quad y \geq 0,\ \alpha \in (0,1),$$

and note that, from (24)–(25), $L_\alpha(K_\alpha) = \inf_{y\geq 0} L_\alpha(y)$ for each $\alpha \in (0,1)$, and also that $L_\alpha(\cdot)\downarrow L(\cdot)$ as $\alpha \uparrow 1$, where $L(\cdot)$ is the function in (18). From these facts, we see that

$$L_\alpha(K^*) \geq L_\alpha(K_\alpha) \geq L(K_\alpha) \geq L(K^*) \quad \forall \alpha \in (0,1),$$

where K^* is as in Remark 4.7(b). Thus, from Remark 4.7(b), we also obtain

$$\varrho^* = L(K^*) = \lim_{\alpha\to 1^-} L_\alpha(K_\alpha) = \lim_{\alpha\to 1^-} (1-\alpha)V_\alpha(0).$$

Therefore, the K^*-threshold policy is strong EAC- and SPAC-optimal. In fact,

$$j^* = \varrho^* = L(K^*) = J(f_{K^*}, x) \quad \forall x \in \mathbf{X}. \ \blacksquare$$

Remark 4.9. In Vega-Amaya and Montes-de-Oca (1997) the EAC-optimal control problem with the one-step cost function (17) is solved using the vanishing discount factor approach and, instead of Assumption 4.1(b), the following:

Assumption 4.1(b′). The demand variable w_0 has a bounded continuous density function.

In that paper it is shown, under Assumptions 4.1(a) and 4.1(b′), that

$$J(f_{K^*}, x) = \varrho^* = \lim_{\alpha\to 1^-} (1-\alpha)V_\alpha(x) \quad \forall x \in \mathbf{X}.$$

Thus, proceeding as in the proof of Theorem 4.8, one can conclude that $\varrho^* = j^*$ and f_{K^*} is strong EAC-optimal and SPAC-optimal.

Example C. An alternative to measure the inventory system performance is to consider quadratic holding and production costs, that is,

$$C(x,a) = R(x - \overline{x})^2 + S(a - \overline{a})^2, \quad (x,a) \in \mathbf{K}, \tag{26}$$

where R and S are positive constants, and $\overline{x} \in \mathbf{X}$ and $\overline{a} \in \mathbf{A}$ denote the target inventory and production levels, respectively. We now suppose:

ASSUMPTION 4.10. The second moment of the demand variables is finite, that is, $\int_0^\infty y^2\, G(dy) < \infty$.

For the cost function (26), Assumption 2.1(a)–(b) trivially holds, while Assumption 4.10 ensures that j^* is finite. Indeed, consider the stationary policy $f(x) = 0$, $x \in \mathbf{X}$, and compute its average cost to obtain

$$J(f, x) = \overline{x}^2 + \overline{a}^2 \qquad \forall x \in \mathbf{X}.$$

These facts yield the following result:

THEOREM 4.11. *Suppose that Assumptions* 4.1 *and* 4.10 *hold. Then there exists a positive Harris recurrent policy $\varphi^* \in \Phi_{\mathrm{P}}$ which is SPAC-optimal.*

EXAMPLE D. Parlar and Rempa/la (1992) study a finite horizon control problem for an inventory system considering a variant of (26), in which there is a "cost free interval" containing the target stock level. More precisely, they take as the holding cost the function

$$\overline{C}(y) := \begin{cases} R_1(y-\alpha)^2 & \text{if } 0 \le y < \alpha, \\ 0 & \text{if } \alpha \le y \le \beta, \\ R_2(y-\beta)^2 & \text{if } y > \beta, \end{cases}$$

where $0 < \alpha < \beta$ and R_1, R_2 are positive constants, and the one-step cost function is given as

$$C(x, a) = E\overline{C}(x + a - w_0) + S(a - \overline{a})^2, \qquad (x, a) \in \mathbf{K}. \tag{27}$$

As in Example C, it is easy to establish the following results.

THEOREM 4.12. *Suppose that Assumptions* 4.1 *and* 4.10 *hold. Then there exists $\varphi^* \in \Phi_{\mathrm{P}}$ which is SPAC-optimal.*

5. Proof of Theorems 3.4 and 3.5. Before the proofs, we introduce some notation and preliminary results, including a useful lemma concerning a class of "approximating" functions.

Let $(Y, \mathcal{T})$ be a separable metrizable space. Denote by $\mathcal{C}_{\mathrm{b}}(Y)$ the space of continuous bounded functions defined on Y with the supremum norm. For each metric d on Y, $\mathcal{U}_d(Y)$ stands for the class of functions in $\mathcal{C}_{\mathrm{b}}(Y)$ which are uniformly continuous with respect to d. We take $\mathcal{U}_d(Y)$ to have the relative topology of $\mathcal{C}_{\mathrm{b}}(Y)$.

The following lemma has an important role in the proof of Theorem 3.4.

LEMMA 5.1. *Let $(Y, \mathcal{T})$ be a separable metrizable space. Then there exists a metric d^* on Y consistent with $\mathcal{T}$ such that*:

(a) *the subspace $\mathcal{U}_{d^*}(Y)$ is separable*;

(b) *for each $u \in \mathcal{C}_{\mathrm{b}}(Y)$ there exist sequences $\{v_n^0\}$ and $\{v_n^1\}$ in $\mathcal{U}_{d^*}(Y)$ such that $v_n^0 \uparrow u$ and $v_n^0 \downarrow u$ as $n \to \infty$.*

The proof of Lemma 5.1 is given in Bertsekas and Shreve (1978) [see Corollary 7.6.1, Proposition 7.9 and Lemma 7.7, on pp. 113, 116 and 125, respectively].

Lemma 5.2. *Let X and Y be Borel spaces and γ a probability measure on $X \times Y$. Then there exist a stochastic kernel $\varphi(\cdot \mid \cdot)$ on Y given X and a measure μ on X such that*

$$\gamma(B \times D) = \int\limits_B \varphi(D \mid x)\ \mu(dy) \quad \forall D \in \mathcal{B}(Y),\ B \in \mathcal{B}(X); \tag{28}$$

hence,

$$\mu(B) = \gamma(B \times Y) \quad \forall B \in \mathcal{B}(X).$$

The measure μ in (28) is called the *marginal distribution* or *projection measure* of γ on X. For the proof of this result see, for instance, Bertsekas and Shreve (1978), Corollary 7.27.2, p. 139, or Hinderer (1970), Theorem 2, p. 189.

Remark 5.3. Let ν and $\nu_n,\ n \in \mathbb{N}$, be measures on $X \times Y$ and denote by μ and $\mu_n,\ n \in \mathbb{N}$, the corresponding marginal distributions. It is easy to verify that if $\{\nu_n\}$ converges weakly to ν, then $\{\mu_n\}$ converges weakly to μ.

We now proceed to prove Theorem 3.4.

Proof of Theorem 3.4. Let $\delta \in \Delta$ and $\nu \in \mathbf{P}(\mathbf{X})$ be arbitrary but fixed and define the random variable

$$\widehat{J} := \liminf_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t, a_t).$$

Observe that if for some realization of the process $\{(x_t, a_t)\}$ generated by δ and ν we have $\widehat{J} = \infty$, then the assertion in Theorem 3.4 trivially holds. Thus, we can assume without loss of generality that $\widehat{J}$ is a finite random variable. Now define the *empirical measures*

$$\gamma_n(\Gamma) := \frac{1}{n} \sum_{t=0}^{n-1} \mathbf{I}_\Gamma(x_t, a_t), \qquad \Gamma \in \mathcal{B}(\mathbf{X} \times \mathbf{A}),\ n \ge 1,$$

where $\mathbf{I}_\Gamma(\cdot)$ denotes the indicator function of Γ. Observe that the measures $\{\gamma_n(\cdot)\}$ are concentrated on $\mathbf{K}$ and also that

$$\infty > \widehat{J} = \liminf_{n\to\infty} \int\limits_{\mathbf{K}} C(x,a)\, \gamma_n(d(x,a)).$$

The proof is divided into two parts. In the first one, we prove that for each $\omega \in \Omega$, there exists a measure $\gamma_\omega(\cdot) \in \mathbf{P}(\mathbf{K})$ such that

$$\widehat{J}(\omega) \ge \int\limits_{\mathbf{K}} C(x,a)\, \gamma^\omega(d(x,a)). \tag{29}$$

Thus, decomposing the measure $\gamma^\omega(\cdot)$ (see Lemma 5.2) as

$$\gamma^\omega(B\times D)=\int_B \varphi^\omega(D\,|\,x)\,\mu^\omega(dx),\qquad B\times D\in\mathcal{B}(\mathbf{X}\times\mathbf{A}),\tag{30}$$

where $\varphi^\omega\in\mathbf{P}(\mathbf{A}\,|\,\mathbf{X})$ and $\mu^\omega\in\mathbf{P}(\mathbf{X})$, we obtain

$$\widehat{J}(\omega)\ge\int_{\mathbf{K}} C(x,\varphi^\omega)\,\mu^\omega(dx).\tag{31}$$

In the second part, we prove that (P_ν^δ-almost surely) $\mu^\omega(\cdot)$ is an invariant probability measure for the transition law $Q(\cdot\,|\,\cdot,\varphi^\omega)$, that is, $\varphi^\omega(\cdot\,|\,\cdot)$ is a relaxed *stable* policy. From this and (31), we conclude that

$$\widehat{J}(\omega)\ge J(\varphi^\omega,\mu^\omega)\ge j^*.\tag{32}$$

PART 1. Fix $\omega\in\Omega$, and choose a sequence $\{n_k\}$ such that

$$\widehat{J}(\omega)=\lim_{k\to\infty}\int_{\mathbf{K}} C(x,a)\,\gamma^\omega_{n_k}(d(x,a));$$

thus,

$$\sup_k\int_{\mathbf{K}} C(x,a)\,\gamma^\omega_{n_k}(d(x,a))<\infty.$$

From Assumption 2.1(b), the latter fact is equivalent to the tightness of the sequence $\{\gamma^\omega_{n_k}(\cdot)\}$ [Meyn and Tweedie (1993), Lemma D.5.3(i)]. Thus, by Prokhorov's Theorem [Billingsley (1968), p. 37], we can pick a subsequence $\{m_k\}$ such that $\{\gamma_{m_k}(\cdot)\}$ converges weakly to a probability measure $\gamma^\omega(\cdot)\in\mathbf{P}(\mathbf{K})$, that is,

$$\int_{\mathbf{K}} v(x,a)\,\gamma^\omega_{m_k}(d(x,a))\to\int_{\mathbf{K}} v(x,a)\,\gamma^\omega(d(x,a))\qquad\forall v\in\mathcal{C}_{\mathrm{b}}(\mathbf{K}).\tag{33}$$

From this and Assumption 2.1(a), we obtain (29); hence, using (30), we conclude that (31) holds.

PART 2. Let d^* be as in Lemma 5.1 and $\mathcal{U}$ a countable dense subset of $\mathcal{U}_{d^*}(\mathbf{X})$ [see Lemma 5.1(a)]. Define, for each $u\in\mathcal{U}$, the function

$$Lu(x,a):=\int_{\mathbf{X}} u(y)\,Q(dy\,|\,x,a)-u(x),\qquad (x,a)\in\mathbf{K},$$

and also the process

$$\begin{aligned}&M_0(u):=u(x_0),\\ &M_n(u):=u(x_n)-\sum_{t=0}^{n-1}Lu(x_t,a_t),\qquad n\ge1.\end{aligned}\tag{34}$$

Observe that for each $u\in\mathcal{U}$, $Lu\in\mathcal{C}_{\mathrm{b}}(\mathbf{K})$ and also that $\{M_n(u)\}$ is a P_ν^δ-martingale with respect to the filtration $\{\sigma(h_n,a_n)\}$. Then the Law of Large Numbers for martingales [Hall and Heyde (1980), Theorem 2.18] yields that for each $u \in \mathcal{U}$ there exists a measurable subset U_u of Ω such that $P_\nu^\delta(U_u) = 1$ and

$$\lim_{n\to\infty} \frac{1}{n} M_n(u) = 0 \quad \text{on } U_u,$$

which implies that

$$\lim_{n\to\infty} \int_{\mathbf{K}} Lu(x,a)\,\gamma_n^\omega(d(x,a)) = 0 \quad \forall \omega \in U_u.$$

Then

$$\lim_{n\to\infty} \int_{\mathbf{K}} Lu(x,a)\,\gamma_n^\omega(d(x,a)) = 0 \quad \forall u \in \mathcal{U} \text{ and } \omega \in U := \bigcap_{u\in\mathcal{U}} U_u.$$

Next, for each $\omega \in U$, choose a sequence $\{m_k\} = \{m_k(\omega)\}$ as in (33). Thus,

$$\int_{\mathbf{K}} Lu(x,a)\,\gamma^\omega(d(x,a)) = 0 \quad \forall u \in \mathcal{U},$$

Hence, using the fact that L is a difference of two monotonic operators and standard "limit" arguments, from Lemma 5.1(b) we see that

$$\int_{\mathbf{K}} Lu(x,a)\,\gamma^\omega(d(x,a)) = 0 \quad \forall u \in \mathcal{C}_{\mathrm{b}}(\mathbf{X}),$$

which yields, after decomposing the measure $\gamma^\omega(\cdot)$ as in (30),

$$\int_{\mathbf{X}} u(x)\,\mu^\omega(dx) = \int_{\mathbf{X}}\int_{\mathbf{X}} u(y)\,Q(dy\,|\,x,\varphi^\omega)\,\mu^\omega(dx) \quad \forall u \in \mathcal{C}_{\mathrm{b}}(\mathbf{X}).$$

This implies that $\mu^\omega(\cdot)$ is an invariant probability measure for $Q(\cdot\,|\,\cdot,\varphi^\omega)$. Finally, combining this fact with (31), we conclude that

$$\widehat{J}(\omega) \ge J(\varphi^\omega,\mu^\omega) \ge j^* \quad \forall \omega \in U,$$

which completes the proof, since the subset U has probability one with respect to P_ν^δ. ∎

Proof of Theorem 3.5. (a) To prove this part, note that it only remains to show that any EAC-optimal policy is strong EAC-optimal. Thus, suppose that δ^* is EAC-optimal. Now observe, from Theorem 3.4 and Fatou's Lemma, that

$$J(\delta,x) \ge \liminf_{n\to\infty} \frac{1}{n} E_x^\delta \sum_{t=0}^{n-1} C(x_t,a_t) \ge j^* \quad \forall \delta \in \Delta,\ x \in \mathbf{X}. \tag{35}$$

Then, putting $\delta = \delta^*$ in (35), we have

$$J(\delta^*,x) = \lim_{n\to\infty} \frac{1}{n} E_x^{\delta^*} \sum_{t=0}^{n-1} C(x_t,a_t) = j^* \quad \forall x \in \mathbf{X},$$

which combined with (35) proves that δ^* is strong EAC-optimal.

(b) Suppose that (δ, ν) is a minimum pair, i.e., $J(\delta, \nu) = j^*$. Then, from Theorem 3.4 and Fatou's Lemma, we see that

$$j^* = J(\delta, \nu) \geq \liminf_{n\to\infty} \frac{1}{n} E_\nu^\delta \sum_{t=0}^{n-1} C(x_t, a_t) \geq E_\nu^\delta \Big[\liminf_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t, a_t) \Big] \geq j^*;$$

hence,

$$E_\nu^\delta \Big[\liminf_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t, a_t) \Big] = j^*,$$

which, jointly with Theorem 3.4, implies that

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t, a_t) = j^* \qquad P_\nu^\delta\text{-almost surely.}$$

(c) Let $\varphi \in \Phi_{\mathrm{S}}$ and μ_φ an associated invariant probability measure. The Individual Ergodic Theorem [Hernández-Lerma and Lasserre (1996), Theorem E.13, p. 189; Dudley (1989), Theorem 8.4.1, p. 209] yields

$$J(\varphi, x) = \lim_{n\to\infty} \frac{1}{n} E_x^\varphi \sum_{t=0}^{n-1} C(x_t, a_t) \qquad \text{for } \mu_\varphi\text{-almost all } x \in \mathbf{X}, \tag{36}$$

and

$$\int_{\mathbf{X}} J(\varphi, x)\, \mu_\varphi(dx) = \int_{\mathbf{X}} C(x, \varphi)\, \mu_\varphi(dx). \tag{37}$$

Suppose that (φ, μ_φ) is a minimum pair. Then, from (37),

$$j^* = \int_{\mathbf{X}} J(\varphi, x)\, \mu_\varphi(dx).$$

Next consider the set $B := \{x \in \mathbf{X} : J(\varphi, x) > j^*\}$ and observe that $j^* \mu_\varphi(B) = \int_B J(\varphi, x)\, \mu_\varphi(dx)$, which implies that $\mu_\varphi(B) = 0$, i.e., $J(\varphi, x) = j^*$ for μ_φ-almost all $x \in \mathbf{X}$.

Now suppose that $J(\varphi, x) = j^*$ for μ_φ-almost all $x \in \mathbf{X}$. Then, from (37), we see that (φ, μ_φ) is a minimum pair. ■

6. Proof of Theorem 3.6. For the proof of Theorem 3.6 we require some preliminary results which are collected in Remarks 6.1 and 6.2.

REMARK 6.1. (a) Let $\delta \in \Delta$, $\nu \in \mathbf{P}(\mathbf{X})$ and $\alpha \in (0, 1)$ be fixed but arbitrary. Define

$$\gamma(\Gamma) := (1 - \alpha) \sum_{t=0}^{n-1} \alpha^t P_\nu^\delta[(x_t, a_t) \in \Gamma], \qquad \Gamma \in \mathcal{B}(\mathbf{X} \times \mathbf{A}). \tag{38}$$

Observe that $\gamma(\cdot)$ is a probability measure on $\mathbf{X} \times \mathbf{A}$ and it is concentrated on $\mathbf{K}$. Moreover, for any measurable function v on $\mathbf{K}$,

$$\int_{\mathbf{K}} v(x,a)\,\gamma(d(x,a)) = (1-\alpha)\sum_{t=0}^{\infty}\alpha^t E_\nu^\delta C(x_t,a_t); \tag{39}$$

in particular,

$$\int_{\mathbf{K}} C(x,a)\,\gamma(d(x,a)) = (1-\alpha)V_\alpha(\delta,\nu). \tag{40}$$

(b) Denote by $\mu(\cdot)$ the marginal distribution of $\gamma(\cdot)$, that is,

$$\mu(B) := \gamma(B\times\mathbf{A}), \qquad B\in\mathcal{B}(\mathbf{X}).$$

One can check that the measures $\mu(\cdot),\gamma(\cdot)$ and $\nu(\cdot)$ satisfy the following "discounted equation" [Hernández-Lerma and Lasserre (1996), Remark 6.3.1, p. 133]:

$$\mu(B) = (1-\alpha)\nu(B) + \alpha\int_{\mathbf{K}} Q(B\,|\,x,a)\,\gamma(d(x,a)) \qquad \forall B\in\mathcal{B}(\mathbf{X}). \tag{41}$$

Remark 6.2. (a) Define

$$\varrho := \liminf_{\alpha\to1^-}(1-\alpha)m_\alpha.$$

From a well-known Abelian Theorem [Hernández-Lerma and Lasserre (1996), Lemma 5.3.1, p. 84], we have

$$\varrho \le \limsup_{\alpha\to1^-}(1-\alpha)m_\alpha \le j^*. \tag{42}$$

(b) For each $\varepsilon>0$ and $\alpha\in(0,1)$, there exist $\delta_\alpha\in\Delta$ and $\nu_\alpha\in\mathbf{P}(\mathbf{X})$ such that $V_\alpha(\delta_\alpha,\nu_\alpha)\le m_\alpha+\varepsilon$. Thus,

$$\varrho = \liminf_{\alpha\to1^-}(1-\alpha)m_\alpha = \liminf_{\alpha\to1^-}(1-\alpha)V_\alpha(\delta_\alpha,\nu_\alpha). \tag{43}$$

Proof of Theorem 3.6. (a) By (43) we can pick a sequence $\{(\delta_{\alpha(n)},\nu_{\alpha(n)})\}$ such that

$$\varrho = \lim_{n\to\infty}(1-\alpha(n))V_{\alpha(n)}(\delta_{\alpha(n)},\nu_{\alpha(n)}). \tag{44}$$

Now, for each $n\in\mathbb{N}$, define

$$\gamma_n(\Gamma) := (1-\alpha(n))\sum_{t=0}^{\infty}[\alpha(n)]^t P_{\nu_{\alpha(n)}}^{\delta_{\alpha(n)}}[(x_t,a_t)\in\Gamma], \qquad \Gamma\in\mathcal{B}(\mathbf{X}\times\mathbf{A}).$$

Next, from (40), observe that

$$(1-\alpha(n))V_{\alpha(n)}(\delta_{\alpha(n)},\nu_{\alpha(n)}) = \int_{\mathbf{K}} C(x,a)\,\gamma_n(d(x,a)).$$

Thus, from (42) and (44),

$$\varrho = \lim_{n\to\infty}\int_{\mathbf{K}} C(x,a)\,\gamma_n(d(x,a)) \le j^* < \infty,$$

which implies that

$$\sup_n \int_{\mathbf{K}} C(x,a)\,\gamma_n(d(x,a)) < \infty.$$

Then the sequence $\{\gamma_n(\cdot)\}$ of measures is tight. Hence, by Prokhorov's Theorem, there exists a subsequence of $\{\gamma_n(\cdot)\}$, which we denote again by $\{\gamma_n(\cdot)\}$ to avoid cumbersome notation, that converges weakly to a probability measure $\gamma^*(\cdot) \in \mathbf{P}(\mathbf{X})$, that is,

$$\int_{\mathbf{K}} v(x,a)\,\gamma_n(d(x,a)) \to \int_{\mathbf{K}} v(x,a)\,\gamma^*(d(x,a)) \quad \forall v \in \mathcal{C}_{\mathrm{b}}(\mathbf{K}). \tag{45}$$

Thus, since $C(\cdot,\cdot)$ is lower semicontinuous on $\mathbf{K}$, we have

$$j^* \geq \lim_{n\to\infty} \int_{\mathbf{K}} C(x,a)\,\gamma_n(d(x,a)) \geq \int_{\mathbf{K}} C(x,a)\,\gamma^*(d(x,a)). \tag{46}$$

We shall prove in the following that there exists a relaxed stable policy φ^* with invariant probability measure $\mu^*(\cdot)$ such that

$$\int_{\mathbf{K}} C(x,a)\,\gamma^*(d(x,a)) = \int_{\mathbf{X}} C_{\varphi^*}(x)\,\mu^*(dx) = J(\varphi^*,\mu^*), \tag{47}$$

from which, combined with (46), we conclude that

$$j^* = \lim_{\alpha\to 1^-} (1-\alpha)m_\alpha = J(\varphi^*,\mu^*).$$

To prove (47), first note, from Lemma 5.2, that there exist relaxed policies (or stochastic kernels on $\mathbf{A}$ given $\mathbf{X}$) φ_n, φ^* and measures $\mu_n, \mu \in \mathbf{P}(\mathbf{X})$ such that for all $B \times D \in \mathcal{B}(\mathbf{X}\times\mathbf{A})$ and $n \in \mathbf{N}$,

$$\gamma_n(B\times D) = \int_B \varphi_n(D\,|\,x)\,\mu_n(dx) \quad \text{and} \quad \gamma^*(B\times D) = \int_B \varphi^*(D\,|\,x)\,\mu^*(dx).$$

Moreover, the weak convergence of $\{\gamma_n(\cdot)\}$ to $\gamma^*(\cdot)$ implies (see Remark 5.3) the weak convergence of $\{\mu_n(\cdot)\}$ to $\mu^*(\cdot)$, that is,

$$\int_{\mathbf{X}} v(x)\,\mu_n(dx) \to \int_{\mathbf{X}} v(x)\,\mu^*(dx) \quad \forall v \in \mathcal{C}_{\mathrm{b}}(\mathbf{X}). \tag{48}$$

On the other hand, from Remark 6.1(b),

$$\mu_n(B) = (1-\alpha(n))\nu_{\alpha(n)}(B) + \alpha(n)\int_{\mathbf{K}} Q(B\,|\,x,a)\,\gamma_n(d(x,a)) \quad \forall B \in \mathcal{B}(\mathbf{X}),$$

which implies

$$\begin{aligned}\int_{\mathbf{X}} v(x)\mu_n(dx) &= (1-\alpha(n))\int_{\mathbf{X}} v(x)\,\nu_{\alpha(n)}(dx)\\ &\quad + \alpha(n)\int_{\mathbf{K}}\int_{\mathbf{X}} v(y)\,Q(dy\,|\,x,a)\,\gamma_n(d(x,a))\end{aligned} \tag{49}$$

for all $v \in \mathcal{C}_{\mathrm{b}}(\mathbf{X})$.

Now observe that for each $v \in \mathcal{C}_{\mathrm{b}}(\mathbf{X})$, the sequence $\int_{\mathbf{X}} v(x)\,\nu_{\alpha(n)}(dx)$, $n \in \mathbb{N}$, is bounded, and also that the function $\int_{\mathbf{X}} v(y)\,Q(dy \mid \cdot,\cdot)$ is in $\mathcal{C}_{\mathrm{b}}(\mathbf{K})$. Thus, from (45) and (48), letting n go to infinity in (49) we obtain

$$\int_{\mathbf{X}} v(x)\,\mu^*(dx) = \int_{\mathbf{K}}\int_{\mathbf{X}} v(y)\,Q(dy \mid x,a)\,\gamma^*(d(x,a)) \qquad \forall v \in \mathcal{C}_{\mathrm{b}}(\mathbf{X}),$$

which is equivalent to

$$\int_{\mathbf{X}} v(x)\,\mu^*(dx) = \int_{\mathbf{X}}\int_{\mathbf{X}} v(y)\,Q(dy \mid x,\varphi^*)\,\mu^*(dx) \qquad \forall v \in \mathcal{C}_{\mathrm{b}}(\mathbf{X}).$$

Then $\mu^*(\cdot)$ is an invariant probability measure for the transition probability $Q(\cdot \mid \cdot,\varphi^*)$, that is, φ^* is a stable policy. Hence, (47) holds, that is,

$$\int_{\mathbf{K}} C(x,a)\,\gamma^*(d(x,a)) = \int_{\mathbf{X}} C_\varphi(x)\,\mu^*(dx) = J(\varphi^*,\mu^*) \geq j^*.$$

Therefore, $j^* = J(\varphi^*,\mu^*) = \lim_{\alpha\to 1^-}(1-\alpha)m_\alpha$.

(b) Suppose that the policy φ^* in (a) is positive Harris recurrent. Thus, by the Law of Large Numbers for Markov chains [Meyn and Tweedie (1993), Theorem 17.01, p. 411], for all initial distributions $\nu \in \mathbf{P}(\mathbf{X})$,

$$J_0(\varphi^*,\nu) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=0}^{n-1} C(x_t,a_t) = j^* \qquad P_\nu^{\varphi^*}\text{-almost surely.}$$

This and Theorem 3.4 show that φ^* is SPAC-optimal. ■

Acknowledgments. The author wishes to thank Professors Onésimo Hernández-Lerma and /Lukasz Stettner for many useful comments on this paper.

References

A. Arapostathis *et al.* (1993), *Discrete time controlled Markov processes with an average cost criterion: A survey*, SIAM J. Control Optim. 31, 282–344.

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

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

P. Billingsley (1968), *Convergence of Probability Measures*, Wiley.

V. S. Borkar (1991), *Topics in Controlled Markov Chains*, Pitman Res. Notes Math. Ser. 240, Longman Sci. Tech.

R. Cavazos-Cadena and E. Fernández-Gaucherand (1995), *Denumerable controlled Markov chains with average reward criterion: sample path optimality*, Z. Oper. Res. 41, 89–108.

R. M. Dudley (1989), *Real Analysis and Probability*, Wadsworth & Brooks.

P. Hall and C. C. Heyde (1980), *Martingale Limit Theory and Its Application*, Academic Press.

O. Hernández-Lerma (1993), *Existence of average optimal policies in Markov control processes with strictly unbounded costs*, Kybernetika 29, 1–17.

O. Hernández-Lerma and J. B. Lasserre (1995), *Invariant probabilities for Feller–Markov chains*, J. Appl. Math. Stochastic Anal. 8, 341–345.

O. Hernández-Lerma and J. B. Lasserre (1996), *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer, New York.

O. Hernández-Lerma and J. B. Lasserre (1997), *Policy iteration for average cost Markov control processes on Borel spaces*, Acta Appl. Math., to appear.

O. Hernández-Lerma and M. Muñoz-de-Osak (1992), *Discrete-time Markov control processes with discounted unbounded cost: optimality criteria* Kybernetika 28, 191–212.

O. Hernández-Lerma, O. Vega-Amaya and G. Carrasco (1998), *Sample-path optimality and variance-minimization of average cost Markov control processes*, Reporte Interno #236, Departamento de Matemáticas, CINVESTAV-IPN, México City.

K. Hinderer (1970), *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameters*, Lecture Notes in Oper. Res. and Math. Systems 33, Springer, Berlin.

J. B. Lasserre (1997), *Sample-path average optimality for Markov control processes*, Report No. 97102, LAAS-CNRS, Toulouse.

H. L. Lee and S. Nahmias (1993), *Single-product, single-location models*, in: *Logistic of Production and Inventory*, S. C. Graves, A. H. G. Rinnooy Kan and P. H. Zipkin (eds.), Handbooks in Operations Research and Management Science, Vol. 4, North-Holland, 3–51.

P. Mandl and M. Lausmanová (1991), *Two extensions of asymptotic methods in controlled Markov chains*, Ann. Oper. Res. 28, 67–80.

S. P. Meyn (1989), *Ergodic theorems for discrete time stochastic systems using a stochastic Lyapunov function*, SIAM J. Control Optim. 27, 1409–1439.

S. P. Meyn (1995), *The policy iteration algorithm for average reward Markov decision processes with general state space*, preprint, Coordinated Science Laboratory, University of Illinois, Urbana, IL.

S. P. Meyn and R. L. Tweedie (1993), *Markov Chains and Stochastic Stability*, Springer, London.

M. Parlar and R. Rempała (1992), *Stochastic inventory problem with piecewise quadratic holding cost function containing a cost-free interval*, J. Optim. Theory Appl. 75, 133–153.

O. Vega-Amaya and R. Montes-de-Oca (1998), *Application of average dynamic programming to inventory systems*, Math. Methods Oper. Res. 47, 451–471.

Oscar Vega-Amaya
Departamento de Matemáticas
Universidad de Sonora
Blvd. Transversal y Rosales s/n
C.P. 83000
Hermosillo, Sonora, México
E-mail: ovega@fisica.uson.mx

Received on 7.4.1997;
revised version on 15.12.1998