APPLICATIONES MATHEMATICAE
27,3 (2000), pp. 343–367

O. VEGA-AMAYA and F. LUQUE-VÁSQUEZ (Sonora)

SAMPLE-PATH AVERAGE COST OPTIMALITY FOR SEMI-MARKOV CONTROL PROCESSES ON BOREL SPACES: UNBOUNDED COSTS AND MEAN HOLDING TIMES

Abstract. We deal with semi-Markov control processes (SMCPs) on Borel spaces with unbounded cost and mean holding time. Under suitable growth conditions on the cost function and the mean holding time, together with stability properties of the embedded Markov chains, we show the equivalence of several average cost criteria as well as the existence of stationary optimal policies with respect to each of these criteria.

1. Introduction. A quick glance at the literature dealing with semi-Markov control processes (SMCPs) shows the following facts (see, e.g., [2], [3], [6]–[9], [13], [18], [19], [21], [22], [25], [26], [25]–[29]):

(i) there are several ways to measure the performance of the controlled system by means of a (long-run) average cost;

(ii) most of the papers use *expected* cost criteria, with a small number of exceptions: Bhatnagar and Borkar [2], Kurano [18], [19], who consider different sample-path average cost criteria;

(iii) almost all papers consider either *denumerable* spaces ([2], [6]–[9], [13], [21], [22], [25], [27]–[29]) or Borel spaces and *bounded costs* (see [3], [18], [19]);

(iv) moreover, in all papers it is assumed that the mean holding time is a *bounded function.*

In this paper our main concern is *sample-path average cost* (*SPAC*-) *optimality for SMCPs on Borel spaces*, *with unbounded costs and unbounded*

2000 *Mathematics Subject Classification*: 90C40, 93E20.

Key words and phrases: semi-Markov control processes, sample-path average costs.

The research was partially supported by CONACYT (México) under Grants 28309 E and 3115P-E9608.

mean holding time. We show the existence of a stationary policy which is optimal with respect to four average cost criteria. We begin with an optimality analysis of the ratio expected average cost ([6]–[9]), that is, analyzing the upper limit of the ratios of expected average costs during the first n transitions to the expected time for these n transitions to occur when n goes to infinity [see (5)]. In order to guarantee the existence of a stationary optimal policy for this criterion we suppose, besides a usual continuity/compactness requirement, that the cost function and the mean holding time satisfy a growth condition and also that the embedded Markov chains have suitable stability properties (see Assumptions 3.1, 3.2, 3.4). Then, in Theorem 3.5, we prove the existence of a solution to the *Average Cost Optimality Equation* (*ACOE*), which, in turn, yields the existence of stationary optimal policies. Similar results were previously obtained under weaker recurrence conditions, but in the denumerable space case ([6]–[8], [27]), and also for Borel spaces with bounded costs ([3], [18], [19]). Moreover, our approach to ensure the existence of a solution to the ACOE is a direct one in the sense that we use *neither Schweitzer's data transformation* (see [28]) *nor the vanishing discount factor approach.* Assumptions 3.1, 3.2, 3.4 were previously used ([16], [17], [31]) to study (expected and sample-path) average cost and other undiscounted cost criteria for Markov control processes (MCPs) on Borel spaces and unbounded cost. In fact, our results are extensions to SMCPs of some results in the latter references.

Once the ACOE is established, we study three sample-path average cost criteria. The first one can be thought of as an auxiliary criterion [see (10)–(11)]; the second one is a (direct) sample-path analogue of the ratio expected average cost, that is, it equals the upper limit of the ratios of sample-path costs during the first n transitions to the time for these n transitions occur when n goes to infinity [see (9)]. The third criterion is a time sample-path average cost; more precisely, it is given as the upper limit of ratios of the sample-path costs accumulated in the transitions occurred up to time $t \in (0, \infty)$ to this time when it tends to infinity [see (12)–(13)]. Imposing additional mild growth conditions on the *pth moment*, with $1 < p \leq 2$, of the cost function and the mean holding time, we are able to show that the three criteria coincide almost surely for all policies and initial states and also that a stationary policy is optimal with respect to each one of these sample-path criteria if and only if it is (ratio) expected average cost optimal (see Theorems 3.7, 3.8).

The remainder of the paper is organized as follows. In Section 2 we briefly describe the semi-Markov control model we are interested in and also introduce the average cost optimality criteria. The assumptions and main results are stated in Section 3. The proofs are given in Sections 4, 5 and 6.

We shall use the following terminology and notation throughout the paper. A Borel subset, say X, of a complete and separable metric space is called a *Borel space*, and it is endowed with the Borel σ-algebra $\mathcal{B}(X)$. If X and Y are Borel spaces, a *stochastic kernel on X given Y* is a function $P(\cdot\,|\,\cdot)$ such that $P(\cdot\,|\,y)$ is a probability measure on X for every $y \in Y$ and $P(B\,|\,\cdot)$ is a (Borel-) measurable function on Y for every $B \in \mathcal{B}(X)$. We denote by $\mathbb{N}$ (respectively, $\mathbb{N}_0$) the set of positive (resp., nonnegative) integers; $\mathbb{R}$ (resp., $\mathbb{R}^+$) denotes the set of real (resp., nonnegative) numbers.

2. The optimal control problems. We deal with a *semi-Markov control model* $(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, Q, F, C)$ where the *state space* $\mathbf{X}$ and *action space* $\mathbf{A}$ are Borel spaces. For each $x \in \mathbf{X}$, the subset $A(x)$ of $\mathbf{A}$ is measurable; it is the *admissible control set* for the state x. We also assume that the *admissible pair set*

$$\mathbf{K} := \{(x, a) : x \in \mathbf{X},\ a \in A(x)\}$$

is a Borel subset of $\mathbf{X} \times \mathbf{A}$. We denote by $\mathbf{F}$ the class of measurable functions $f : \mathbf{X} \to \mathbf{A}$ such that $f(x) \in A(x)$ for all $x \in \mathbf{X}$, and we suppose that it is nonempty. Moreover, the *transition law* $Q(B\,|\,x, a)$, with $B \in \mathcal{B}(\mathbf{X})$ and $(x, a) \in \mathbf{K}$, is a stochastic kernel on $\mathbf{X}$ given $\mathbf{K}$, and the *distribution of holding times* $F(t\,|\,x, a)$ is a measurable function on $\mathbf{K}$ for each $t \in \mathbb{R}$ and a distribution function for each $(x, a) \in \mathbf{K}$. Finally, the *one-step cost* C is a measurable function on $\mathbf{K}$.

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

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

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

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

A semi-Markov control model (SMCM) represents a stochastic system evolving as follows: at time $t = 0$ the system is observed in some state $x_0 = x \in \mathbf{X}$ and a control $a_0 = a \in A(x)$ is chosen incurring a cost $C(x, a)$. Then the system remains in the state $x_0 = x$ for a (nonnegative) random time δ_1 with distribution function $F(t\,|\,x, a)$ and jumps to the state $x_1 = y$ according to the probability measure $Q(\cdot\,|\,x, a)$. Immediately after the jump occurs, a control $a_1 = a' \in A(y)$ is chosen and the above process is repeated indefinitely. Thus, for each $n \in \mathbb{N}_0$, we denote by x_n, a_n and δ_{n+1} the state of the system immediately after the nth transition, the control chosen and the corresponding holding (or sojourn) time, respectively. Moreover, we define

$$T_0 := 0, \qquad T_n := T_{n-1} + \delta_n \quad \text{for } n \in \mathbb{N} \tag{3}$$

and

$$\tau(x,a) := \int t\, F(dt \mid x,a), \qquad (x,a) \in \mathbf{K}. \tag{4}$$

Observe that, for each $n \in \mathbb{N}$, T_n is the time at which the nth transition of the system occurs and also that $\tau(x,a)$ is the *mean holding time* in the state x when the control chosen is $a \in A(x)$.

For each $n \in \mathbb{N}_0$, define the space of *admissible histories* until the nth transition by

$$\mathbf{H}_0 := \mathbf{X}, \quad \mathbf{H}_n := (\mathbf{K} \times \mathbb{R}_+)^n \times \mathbf{X} \quad \text{for } n \in \mathbb{N}.$$

DEFINITION 2.1. A *control policy* $\pi = \{\pi_n\}$ is a sequence of stochastic kernels π_n on $\mathbf{A}$ given $\mathbf{H}_n$ satisfying the constraint $\pi_n(A(x_n) \mid h_n) = 1$ for all $h_n = (x_0, a_0, \delta_1, \ldots, x_{n-1}, a_{n-1,}\delta_n, x_n) \in \mathbf{H}_n$, $n \in \mathbb{N}_0$. A policy $\pi = \{\pi_n\}$ is said to be a (deterministic) *stationary policy* if there exists $f \in \mathbf{F}$ such that $\pi_n(\cdot \mid h_n)$ is concentrated at $f(x_n)$ for each $n \in \mathbb{N}_0$. We denote by Π the class of all policies and, following a usual convention, identify the subclass of stationary policies with $\mathbf{F}$.

REMARK 2.2. Let $(\Omega, \mathcal{F})$ be the (canonical) measurable space consisting of the sample space $\Omega := (\mathbf{X} \times \mathbf{A} \times \mathbb{R}_+)^\infty$ and its product σ-algebra. Then, by the Ionescu–Tulcea Theorem ([1], Theorem 2.7.2, p. 109), for each policy $\pi \in \Pi$ and initial state $x_0 = x \in \mathbf{X}$, there exists a probability measure P_x^π on $(\Omega, \mathcal{F})$ satisfying the following: For all $B \in \mathcal{B}(\mathbf{A})$, $C \in \mathcal{B}(\mathbf{X})$, $h_n = (x_0, a_0, \delta_1, \ldots, x_{n-1}, a_{n-1}, \delta_n, x_n) \in \mathbf{H}_n$, $n \in \mathbb{N}_0$, we have

(i) $P_x^\pi[x_0 = x] = 1$;
(ii) $P_x^\pi[a_n \in B \mid h_n] = \pi_n(B \mid h_n)$;
(iii) $P_x^\pi[x_{n+1} \in C \mid h_n, a_n, \delta_{n+1}] = Q(C \mid x_n, a_n)$;
(iv) $P_x^\pi[\delta_{n+1} \le t \mid h_n, a_n] = F(t \mid x_n, a_n)$.

The expectation operator with respect to P_x^π is denoted by E_x^π.

REMARK 2.3. Note that, for an arbitrary policy $\pi \in \Pi$, the distribution of the state of the system x_n may depend on the evolution in the first $n-1$ transitions of the system. However, when a stationary policy $f \in \mathbf{F}$ is used, it follows (from the Markov-like properties in Remark 2.2) that the state process $\{x_n\}$ is a Markov chain with transition probability $Q_f(\cdot \mid x)$. In the latter case, we denote by $Q_f^n(\cdot \mid \cdot)$ the n-step transition probability.

The literature on semi-Markov control processes shows that there are several ways to measure the performance of the systems using an "average cost criterion" (see [2], [9], [18], [19], [21], [22]). In the remainder of this section we introduce the criteria we are concerned with, beginning with the (ratio of) expected average cost: for a policy $\pi \in \Pi$ and initial state

$x_0 = x \in \mathbf{X}$, we define the *expected average cost* (EAC) as

$$J(\pi, x) := \limsup_{n\to\infty} \frac{1}{E_x^\pi T_n} E_x^\pi \sum_{k=0}^{n-1} C(x_k, a_k), \tag{5}$$

and the *optimal expected average cost* function by

$$J^*(x) = \inf_{\pi\in\Pi} J(\pi, x). \tag{6}$$

DEFINITION 2.4. A policy π^* is said to be:

(a) *expected average cost* (*EAC-*) *optimal* if

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

(b) *strong expected average cost* (*strong EAC-*) *optimal* if

$$\liminf_{n\to\infty} \frac{1}{E_x^\pi T_n} E_x^\pi \sum_{k=0}^{n-1} C(x_k, a_k) \geq J(\pi^*, x) \quad \forall x \in \mathbf{X},\ \pi \in \Pi.$$

We shall prove in Theorem 3.5, under suitable continuity/compactness conditions and stability properties, that there exists a stationary strong EAC-optimal policy by showing the existence of a solution to the *Average Cost Optimality Equation.* This problem has been solved in several papers but almost all consider denumerable state space, or general state space but with the assumption that the cost function and the mean holding time are bounded (e.g., for the denumerable case, see [2], [6]–[8], [25]–[27]; and, for the Borel space case, see [3], [18], [19]). Recently, in [23], the existence of a solution to the Average Cost Optimality Equation was proved for the case of Borel spaces and unbounded cost, by assuming that the mean holding time is bounded. Our approach is closely related to that taken in [23], but here we use a weaker stability assumption and do not require that the mean holding time be a bounded function. Moreover, in our proof of the existence of a solution to the Average Cost Optimality Equation we use "direct" arguments, in contrast to those in [23], which are based on the *data transformation* (see [28]) and, implicitly, the *vanishing discount factor approach.*

REMARK 2.5. Note that for $\pi \in \Pi$ and $x \in \mathbf{X}$ we have

$$J(\pi, x) = \limsup_{n\to\infty} \Big[E_x^\pi \sum_{k=0}^{n-1} \tau(x_k, a_k)\Big]^{-1} E_x^\pi \sum_{k=0}^{n-1} C(x_k, a_k). \tag{7}$$

This follows by noting first that the Markov-like properties in Remark 2.2 yield

$$E_x^\pi[T_n \mid h_n, a_n] = \sum_{k=0}^{n-1} \tau(x_k, a_k) \quad \forall h_n \in \mathbf{H}_n,\ a_n \in A(x_n),\ n \in \mathbb{N},$$

then noting that

$$E_x^\pi T_n = E_x^\pi \sum_{k=0}^{n-1} \tau(x_k, a_k) \quad \forall n \in \mathbb{N}, \tag{8}$$

from which we see that (7) holds.

Now, we introduce the sample-path average cost criteria. For a policy $\pi \in \Pi$ and $x \in \mathbf{X}$, define

$$J_0(\pi, x) := \limsup_{n\to\infty} \frac{1}{T_n} \sum_{k=0}^{n-1} C(x_k, a_k), \tag{9}$$

and

$$J_1(\pi, x) := \limsup_{n\to\infty} \frac{1}{\widehat{T}_n} \sum_{k=0}^{n-1} C(x_k, a_k), \tag{10}$$

where

$$\widehat{T}_n := \sum_{k=0}^{n-1} \tau(x_k, a_k), \qquad n \in \mathbb{N}. \tag{11}$$

We also consider a time-average cost criterion defined as follows:

$$J_2(\pi, x) := \limsup_{t\to\infty} \frac{1}{t} \sum_{k=0}^{\eta(t)} C(x_t, a_t), \tag{12}$$

where

$$\eta(t) := \max\{n \geq 0 : T_n \leq t\}, \qquad t \in \mathbb{R}^+. \tag{13}$$

Observe that (9) and (10) are the sample-path analogues of (5) and (7), respectively, and also that (9), (10) and (12) coincide in the Markovian case, that is, when $F(\cdot \mid x, a)$ is concentrated at $t = 1$ for all $(x, a) \in \mathbf{K}$. Indeed, in this case, the expected average cost (5) and the sample-path average costs (9), (10), (12), respectively, become

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

and

$$J_0(\pi, x) = J_1(\pi, x) = J_2(\pi, x) = \limsup_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t, a_t),$$

for all initial states $x \in \mathbf{X}$ and policies $\pi \in \Pi$.

The references dealing with sample-path average cost optimality for unbounded costs are also scarce in the Markovian case; we mention [4], [5] for discrete state spaces, and [17], [20], [30] for Borel spaces.

DEFINITION 2.6. Let $i \in \{0,1,2\}$. A policy π_* is said to be *i-sample path average cost* (*i-SPAC*) *optimal* if there exists a constant ϱ_i such that

$$J_i(\pi_*, x) = \varrho_i \quad P_x^{\pi_*}\text{-a.s.}, \qquad \forall x \in \mathbf{X},$$

and

$$J_i(\pi, x) \geq \varrho_i \quad P_x^{\pi}\text{-a.s.}, \qquad \forall \pi \in \Pi,\ x \in \mathbf{X}.$$

The constant ϱ_i, $i \in \{0,1,2\}$, is called the *i-optimal sample-path average cost.*

We prove in Theorem 3.7, under Assumptions 3.1, 3.2, 3.4 and 3.6, that there exists a stationary 1-SPAC optimal policy f^* and that the 1-optimal sample-path average cost equals the optimal expected average cost function, i.e., $J^*(\cdot) \equiv \varrho_1$. In fact, we prove that a stationary policy $f \in \mathbf{F}$ is EAC-optimal if and only if it is 1-sample-path average cost optimal. Then, under an additional mild assumption, we show in Theorem 3.8 that the sample path average cost criteria (9), (10) and (12) coincide, that is,

$$J_0(\pi, x) = J_1(\pi, x) = J_2(\pi, x) \quad P_x^{\pi}\text{-a.s.}, \qquad \forall \pi \in \Pi,\ x \in \mathbf{X}.$$

Hence, there exists a stationary policy which is optimal with respect to each one of the criteria introduced above and $J^*(\cdot) \equiv \varrho_0 = \varrho_1 = \varrho_2$.

3. Main results and assumptions. To prove the existence of expected and sample-path stationary optimal polices we require to impose suitable conditions on the model. The first one is a combination of standard continuity/compactness requirements together with a growth condition on the cost function C and on the mean holding time τ.

ASSUMPTION 3.1. For each state $x \in \mathbf{X}$:

(a) $A(x)$ is a compact subset of $\mathbf{A}$;
(b) $C(x, \cdot)$ is lower semicontinuous on $A(x)$;
(c) $\tau(x, \cdot)$ is upper semicontinuous on $A(x)$;
(d) $Q(\cdot \mid x, \cdot)$ is strongly continuous on $A(x)$, that is, the mapping

$$a \mapsto \int_{\mathbf{X}} u(y)\, Q(dy \mid x, a)$$

is continuous for each bounded measurable function u on $\mathbf{X}$;

(e) There exist a measurable function $W \geq 1$ on $\mathbf{X}$ and positive constants k_1 and θ such that:

(e1) $|C(x,a)| \leq k_1 W(x)$ and $\theta < \tau(x,a) \leq k_1 W(x)$ $\forall a \in A(x)$;
(e2) $\int_{\mathbf{X}} W(y)\, Q(dy \mid x, \cdot)$ is continuous on $A(x)$.

The following two sets of hypotheses (Assumptions 3.2 and 3.4 below) guarantee that the imbedded Markov chains have suitable stable behavior uniformly in $f \in \mathbf{F}$. These assumptions were already used in a Markovian setting in [16] and [31] to study several undiscounted (expected) cost criteria—such as overtaking optimality, bias optimality and others—and also in [17] to obtain sample-path average cost optimal policies as well as to solve a variance-minimization problem. In fact, Assumption 3.2 is a standard way to obtain (W-) geometric ergodicity for (uncontrolled) Markov chains (see, e.g., [10] or [24], Theorem 16.0.3, p. 383). In what follows we shall use the notation and terminology from these references.

Assumption 3.2. For each stationary policy $f \in \mathbf{F}$:

(a) There exist positive constants $B_f < 1$ and $b_f < \infty$, and a petite subset K_f of $\mathbf{X}$ such that [using the notation (2)]

$$\int_{\mathbf{X}} W(y)\, Q_f(dy \,|\, x) \leq B_f W(x) + b_f \mathbf{I}_{K_f}(x) \qquad \forall x \in \mathbf{X}, \tag{14}$$

where W is the function in Assumption 3.1(e), and $\mathbf{I}_K(\cdot)$ denotes the indicator function of K.

(b) The state process $\{x_n\}$—which under $f \in \mathbf{F}$ is a Markov chain with transition kernel $Q_f(\cdot \,|\, \cdot)$ [see Remark 2.3]—is φ-irreducible and aperiodic, for some (nontrivial) σ-finite measure φ on $\mathbf{X}$, which does not depend on the policy f.

To state some consequences of Assumption 3.2 we need the following notation: let $B_W(\mathbf{X})$ be the linear space of measurable functions u on $\mathbf{X}$ with finite W-*norm*, which is defined as

$$\|u\|_W := \sup_{x \in \mathbf{X}} \frac{|u(x)|}{W(x)}. \tag{15}$$

Moreover, for a measurable function u and measure μ on $\mathbf{X}$, we write

$$\mu(u) := \int_{\mathbf{X}} u(y)\, \mu(dy).$$

Remark 3.3. Under Assumption 3.2, for each stationary policy $f \in \mathbf{F}$ we have:

(a) The Markov chain $\{x_n\}$ induced by f is *positive Harris-recurrent* and its unique invariant probability measure μ_f satisfies $\mu_f(W) < \infty$;

(b) $\{x_n\}$ is W-*geometrically ergodic*; that is, there exist positive constants $\gamma_f < 1$ and $M_f < \infty$ such that

$$\left| \int_{\mathbf{X}} u(y)\, Q_f^n(dy \,|\, x) - \mu_f(u) \right| \leq \|u\|_W M_f \gamma_f^n W(x) \tag{16}$$

for each $u \in B_W(\mathbf{X}), x \in \mathbf{X}$, and $n \in \mathbb{N}_0$.

The proof of Remark 3.3(b) is given in [24], Theorem 16.0.1. Remark 3.3(a) follows from Remark 3.3(b) (see [15], Theorem 3.2), or from [10], Theorem 2.2, which uses a more general Lyapunov condition than (14).

The next assumption concerns the constants γ_f and M_f in (16).

ASSUMPTION 3.4. The constants $M := \sup_f M_f$ and $\gamma := \sup_f \gamma_f$ are such that

$$M < \infty \quad \text{and} \quad \gamma < 1.$$

Assumptions 3.1, 3.2 and 3.4 were previously used in [16] and [17] (see also [31]) to study several undiscounted cost criteria, including the sample-path and expected average cost criteria, for Markov control processes on Borel spaces and unbounded cost. Among these hypotheses, Assumption 3.4 has the inconvenience that it is not imposed directly on the control model. However, [11] provides conditions which guarantee that Assumption 3.4 holds as well as estimates on the constants γ and M. On the other hand, [16] and [23] provide examples (an inventory system and a replacement model, respectively) satisfying Assumptions 3.1, 3.2 and 3.4. In fact, in [23] it is shown that the conditions in [11] imply Assumptions 3.2 and 3.4.

Next, we state our first main result.

THEOREM 3.5. *Suppose that Assumptions* 3.1, 3.2 *and* 3.4 *hold. Then*

(a) *There exist a constant* ϱ^*, *a policy* $f^* \in \mathbf{F}$ *and a function* $h_* \in B_W(\mathbf{X})$ *that solve the* Average Cost Optimality Equation

$$h_*(x) = \min_{a\in A(x)} \Big[C(x,a) - \varrho^*\tau(x,a) + \int_{\mathbf{X}} h_*(y)\, Q(dy\,|\,x,a)\Big] \tag{17}$$

$$= C_{f^*}(x) - \varrho^*\tau_{f^*}(x) + \int_{\mathbf{X}} h_*(y)\, Q_{f^*}(dy\,|\,x) \qquad \forall x \in \mathbf{X}; \tag{18}$$

(b) f^* *is EAC-optimal and*

$$J^*(x) = J(f^*,x) = \varrho^* = \frac{\mu_{f^*}(C_{f^*})}{\mu_{f^*}(\tau_{f^*})} \qquad \forall x \in \mathbf{X};$$

(c) *A policy* $f \in \mathbf{F}$ *is EAC-optimal if and only if it is strong EAC-optimal*; *hence*, f^* *is strong EAC-optimal.*

We have stated in Theorem 3.5 the existence of a strong EAC-optimal stationary policy. In order to show that such a policy is also i-SPAC optimal, $i \in \{0,1,2\}$, we require a suitable strengthening of Assumption 3.1(e1).

ASSUMPTION 3.6. There exist positive constants r_1 and $1 < p \le 2$ such that for all $(x,a) \in \mathbf{K}$,

$$|C(x,a)|^p \le r_1 W(x) \tag{19}$$

and

$$\tau^p(x,a) \leq r_1 W(x) \tag{20}$$

We are now ready to state our second main result.

THEOREM 3.7. *Suppose that Assumptions* 3.1, 3.2, 3.4 *and* 3.6 *are satisfied*, *and let* ϱ^* *be as in Theorem* 3.5. *Then*:

(a) *For each* $\pi \in \Pi$ *and* $x \in \mathbf{X}$,

$$J_1(\pi,x) \geq \liminf_{n\to\infty} \frac{1}{\widehat{T}_n} \sum_{k=0}^{n-1} C(x_k,a_k) \geq \varrho^* \qquad P_x^\pi\text{-a.s.}$$

(b) *A policy* $f \in \mathbf{F}$ *is EAC-optimal if and only if*

$$J_1(f,x) = \varrho^* \qquad P_x^f\text{-a.s.}, \qquad \forall x \in \mathbf{X};$$

hence, *from Theorem* 3.5, *there exists a* 1-*SPAC optimal stationary policy* $f^* \in \mathbf{F}$.

To state our third main result, Theorem 3.8 below, we introduce the function

$$\eta(x,a) := \int_0^\infty t^p\, F(dt \mid x,a), \qquad (x,a) \in \mathbf{K}, \tag{21}$$

where p is as in Assumption 3.6.

THEOREM 3.8. *Suppose that Assumptions* 3.1, 3.2, 3.4 *and* 3.6 *hold*, *and also that there exists a constant* r_2 *such that*

$$\eta(x,a) \leq r_2 W(x) \qquad \forall (x,a) \in \mathbf{K}. \tag{22}$$

Then:

(a) *For each policy* $\pi \in \Pi$ *and state* $x \in \mathbf{X}$,

$$J_0(\pi,x) = J_1(\pi,x) = J_2(\pi,x) \qquad P_x^\pi\text{-a.s.};$$

(b) *The following statements are equivalent*:

(i) $f \in \mathbf{F}$ *is EAC-optimal*;
(ii) $f \in \mathbf{F}$ *is strong EAC-optimal*;
(iii) $f \in \mathbf{F}$ *is* 0-*SPAC optimal*;
(iv) $f \in \mathbf{F}$ *is* 1-*SPAC optimal*;
(v) $f \in \mathbf{F}$ *is* 2-*SPAC optimal.*

Consequently, *by Theorem* 3.7(b), *there exists a stationary policy* f^* *which satisfies each one of conditions* (i)–(v), *and* $\varrho^* = \varrho_0 = \varrho_1 = \varrho_2$.

4. Proof of Theorem 3.5.

For this proof we need some preliminary results, which are collected in Remark 4.1 and Lemmas 4.2–4.4. *Throughout this section we suppose that the assumptions of Theorem* 3.5 *hold.*

REMARK 4.1. (a) Assumptions 3.1(a) and (e2), and the well known Measurable Selection Theorem (see, for instance, [14], Proposition D.5, p. 182) imply the existence of a policy $g \in \mathbf{F}$ such that

$$\sup_{a\in A(x)} \int_{\mathbf{X}} W(y)\, Q(dy \,|\, x,a) = \int_{\mathbf{X}} W(y)\, Q_g(dy \,|\, x) \qquad \forall x \in \mathbf{X}. \tag{23}$$

Then, using (14), we see that for all $x \in \mathbf{X}$ and $f \in \mathbf{F}$,

$$\int_{\mathbf{X}} W(y)\, Q_f(dy \,|\, x) \le B_g W(x) + b_g \mathbf{I}_{K_g}(x), \tag{24}$$

which implies that

$$\mu_f(W) \le b_g/(1-B_g) \qquad \forall f \in \mathbf{F}.$$

Therefore

$$\sup_{f\in\mathbf{F}} \mu_f(W) \le b_g/(1-B_g). \tag{25}$$

(b) Note that, by Assumption 3.1(e1), the constants

$$\varrho_f := \mu_f(C_f)/\mu_f(\tau_f), \qquad f \in \mathbf{F}, \tag{26}$$

are well defined. Moreover, observing that

$$|\varrho_f| \le \frac{1}{\theta}\mu_f(|C_f|) \le \frac{k_1}{\theta}\mu_f(W),$$

where θ and k_1 are as in Assumption 3.1(e1), from (25) we obtain

$$|\varrho_f| \le L \qquad \forall f \in \mathbf{F}, \tag{27}$$

with $L := k_1 b_g/(\theta(1-B_g))$.

(c) For each $f \in \mathbf{F}$, we have

$$J(f,x) = \lim_{n\to\infty} \frac{1}{E_x^f T_n} E_x^f \sum_{k=0}^{n-1} C_f(x_t) = \varrho_f \qquad \forall x \in \mathbf{X}, \tag{28}$$

and

$$\begin{aligned} J_1(f,x) &= \lim_{n\to\infty} \Big[\sum_{k=0}^{n-1} \tau_f(x_k)\Big]^{-1} \sum_{k=0}^{n-1} C_f(x_k) \\ &= \varrho_f \quad P_x^f\text{-a.s,} \quad \forall x \in \mathbf{X}. \end{aligned} \tag{29}$$

To obtain (28), observe that (16) implies that for every $x \in \mathbf{X}$,

$$\lim_{n\to\infty} \frac{1}{n} E_x^f \sum_{k=0}^{n-1} C_f(x_t) = \mu_f(C_f) \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n} E_x^f \sum_{k=0}^{n-1} \tau_f(x_t) = \mu_f(\tau_f).$$

Thus, from (8), we see that (28) holds. Similarly, (29) follows on noting that Remark 3.3(a) and the Strong Law of Large Numbers for Markov chains

([24], p. 411) yield

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} C_f(x_t)=\mu_f(C_f),\quad \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\tau_f(x_t)=\mu_f(\tau_f),\qquad P_x^f\text{-a.s.}$$

Lemma 4.2. *For each $f\in\mathbf{F}$, the function*

$$h_f(x):=\sum_{k=0}^{\infty}E_x^f[C_f(x_k)-\varrho_f\tau_f(x_k)],\qquad x\in\mathbf{X}, \tag{30}$$

is in $B_W(\mathbf{X})$ and is such that the pair (ϱ_f,h_f), with ϱ_f as in (26), *satisfies the Poisson equation*

$$h_f(x)=C_f(x)-\varrho_f\tau_f(x)+\int_{\mathbf{X}} h_f(y)\,Q_f(dy\,|\,x)\qquad\forall x\in\mathbf{X}. \tag{31}$$

Moreover,

$$\sup_{f\in\mathbf{F}}\|h_f\|_W<\infty. \tag{32}$$

Proof. Fix $f\in\mathbf{F}$ and observe that for each $k=0,1,\ldots,$

$$C_f(x_k)-\varrho_f\tau_f(x_k)=[C_f(x_k)-\mu_f(C_f)]+[\varrho_f\mu_f(\tau_f)-\varrho_f\tau_f(x_k)];$$

thus, for each $x\in\mathbf{X}$, we see that

$$|h_f(x)|\le\sum_{k=0}^{\infty}\{|E_x^fC_f(x_k)-\mu_f(C_f)|+|\varrho_fE_x^f\tau_f(x_k)-\varrho_f\mu_f(\tau_f)|\},$$

which, by Remark 3.3(b) and Assumptions 3.1(e1) and 3.4, implies that

$$|h_f(x)|\le M(1+|\varrho_f|)k_1W(x)/(1-\gamma)\qquad\forall x\in\mathbf{X}.$$

Then $h_f\in B_W(\mathbf{X})$ and, by (27), we conclude that (32) holds.

Finally, (31) follows from the Markov property and the fact that

$$h_f(x)=C_f(x)-\varrho_f\tau_f(x)+\sum_{k=1}^{\infty}E_x^f[C_f(x_k)-\varrho_f\tau_f(x_k)] \tag{33}$$

for each $f\in\mathbf{F}$. ∎

Lemma 4.3. *Let $f\in\mathbf{F}$ be a fixed but arbitrary policy. Suppose that there exist a constant ϱ and a function $h\in B_W(\mathbf{X})$ such that*

$$h(x)\ge C_f(x)-\varrho\tau_f(x)+\int_{\mathbf{X}} h(y)\,Q_f(dy\,|\,x)\qquad\forall x\in\mathbf{X}. \tag{34}$$

Then

$$\varrho\ge J(f,\cdot)=\varrho_f. \tag{35}$$

Moreover, if $\varrho = \varrho_f$ *then there exists a measurable subset* N_f *of* $\mathbf{X}$ *with* $\mu_f(N_f) = 1$ *such that*

$$h(x) = h_f(x) + s_f \quad \forall x \in N_f, \qquad h(x) \geq h_f(x) + s_f \quad \forall x \in \mathbf{X}, \tag{36}$$

where $s_f := \mu_f(h - h_f)$.

Proof. The inequality in (35) follows by integration of both sides of (34) with respect to the invariant probability measure μ_f.

Now suppose that $\varrho = \varrho_f$. Subtracting the Poisson equation (31) for f from (34), we find that

$$H(\cdot) := h(\cdot) - h_f(\cdot)$$

is a superharmonic function, that is,

$$H(x) \geq \int_{\mathbf{X}} H(y)\, Q_f(dy \mid x) \quad \forall x \in \mathbf{X},$$

which implies that

$$H(x) \geq \int_{\mathbf{X}} H(y)\, Q_f^n(dy \mid x) \quad \forall x \in \mathbf{X},\ n \in \mathbb{N}.$$

Then, taking the limit as $n \to \infty$, we obtain

$$H(x) \geq \mu_f(H) \quad \forall x \in \mathbf{X}, \tag{37}$$

which implies $\inf_x H(x) = \mu_f(H)$. Thus, $H(\cdot) = \mu_f(H)$ μ_f-a.e., which jointly with (37) proves (36). ∎

LEMMA 4.4. *There exists a policy* f^* *such that*

$$\varrho^* := \inf_f \varrho_f = \varrho_{f^*}.$$

Proof. Consider a sequence $\{f_n\}$ of stationary policies such that $\varrho_n := \varrho_{f_n}$ converges to ϱ^* as $n \to \infty$. Moreover, for each $n \in \mathbb{N}$, let $h_n := h_{f_n} \in B_W(\mathbf{X})$ be the solution to the Poisson equation for the policy f_n as in (30). Now, define

$$h^*(x) := \liminf_{n\to\infty} h_n(x), \qquad x \in \mathbf{X},$$

and observe that, by (32), $h^*(\cdot)$ is in $B_W(\mathbf{X})$.

Next, we shall prove the existence of a policy $f^* \in \mathbf{F}$ such that

$$h^*(x) \geq C_{f^*}(x) - \varrho^* \tau_{f^*}(x) + \int_{\mathbf{X}} h^*(y)\, Q_{f^*}(dy \mid x) \quad \forall x \in \mathbf{X}, \tag{38}$$

which together with Lemma 4.3 implies that $\varrho^* = \varrho_{f^*} \leq \varrho_f$ for all $f \in \mathbf{F}$.

Let $x \in \mathbf{X}$ be a fixed but arbitrary state and $\{n(k)\}$ a subsequence such that

$$h_{n(k)}(x) \to h^*(x) \quad \text{as } k \to \infty.$$

For each $k \in \mathbb{N}$, we have

$$h_{n(k)}(x) = C_{f_{n(k)}}(x) - \varrho_{n(k)}\tau_{f_{n(k)}}(x) + \int_{\mathbf{X}} h_{n(k)}(y)\, Q_{f_{n(k)}}(dy \mid x).$$

Now, since $A(x)$ is a compact subset of $\mathbf{A}$, there exist $a_x \in A(x)$ and a subsequence $\{m(k)\} \subset \{n(k)\}$ such that $f_{m(k)}(x) \to a_x$ as $k \to \infty$. Thus, from Assumption 3.1(a)–(c) and (generalized) Fatou's Lemma (see [1], p. 48), taking lim inf on both sides of the equality

$$h_{m(k)}(x) = C_{f_{m(k)}}(x) - \varrho_{m(k)}\tau_{f_{m(k)}}(x) + \int_{\mathbf{X}} h_{m(k)}(y)\, Q_{f_{m(k)}}(dy \mid x)$$

we obtain

$$h^*(x) \geq C(x, a_x) - \varrho^*\tau(x, a_x) + \int_{\mathbf{X}} [\liminf_{k\to\infty} h_{m(k)}(y)]\, Q(dy \mid x, a_x).$$

Then, since $[\liminf_{k\to\infty} h_{m(k)}(\cdot)] \geq h^*(\cdot)$, we have

$$\begin{aligned} h^*(x) &\geq C(x, a_x) - \varrho^*\tau(x, a_x) + \int_{\mathbf{X}} h^*(y)\, Q(dy \mid x, a_x) \\ &\geq \min_{a\in A(x)} \Big[C(x, a) - \varrho^*\tau(x, a) + \int_{\mathbf{X}} h^*(y)\, Q(dy \mid x, a)\Big]. \end{aligned}$$

Finally, from the Measurable Selection Theorem ([14], Proposition D.5, p. 182), the latter inequality implies the existence of a policy $f^* \in \mathbf{F}$ satisfying (38). ∎

With the above preliminaries we are now ready for the proof of Theorem 3.5 itself. This is based on a "modified" policy iteration algorithm which was already used in [31] and [23] for Markov and semi-Markov control processes, respectively; however, for the sake of completeness, we repeat the arguments here.

Proof of Theorem 3.5. (a) Let $f^* \in \mathbf{F}$ and ϱ^* be as in Lemma 4.4. Put $f_0 := f^*$ and let $h_0 := h_{f_0}$ be a solution to the Poisson equation for f_0 as in (30), that is,

$$h_0(x) = C_{f_0}(x) - \varrho^*\tau_{f_0}(x) + \int_{\mathbf{X}} h_0(y)\, Q_{f_0}(dy \mid x) \qquad \forall x \in \mathbf{X}.$$

Then there exists $f_1 \in \mathbf{F}$ such that for every $x \in \mathbf{X}$,

$$\begin{aligned} h_0(x) &\geq \min_{a\in A(x)} \Big[C(x, a) - \varrho^*\tau(x, a) + \int_{\mathbf{X}} h_0(y)\, Q(dy \mid x, a)\Big] \\ &= C_{f_1}(x) - \varrho^*\tau_{f_1}(x) + \int_{\mathbf{X}} h_0(y)\, Q_{f_1}(dy \mid x). \end{aligned}$$

Now, from Lemma 4.3, we have $\varrho_{f_1} = \varrho^*$ and also there exists a subset $N_1 \subset \mathbf{X}$ with $\mu_{f_1}(N_1) = 1$ such that [writing $h_1 := h_{f_1}$; see (30)]

$$h_0(x) - h_1(x) = s_1 := \mu_{f_1}(h_0 - h_1) \quad \forall x \in N_1$$

and

$$h_0(x) - h_1(x) \geq s_1 \quad \forall x \in \mathbf{X}.$$

Proceeding inductively we obtain sequences $\{f_n\} \subset \mathbf{F}$, $\{h_n\} = \{h_{f_n}\} \subset B_W(\mathbf{X})$, and $\{N_n\} \subset \mathcal{B}(\mathbf{X})$ satisfying the following:

(i) $J(f_n, x) = \varrho_{f_n} = \varrho^*$ for all $x \in \mathbf{X}$, $n \in \mathbb{N}_0$;

(ii) h_n solves the Poisson equation for the policy f_n, that is,

$$h_n(x) = C_{f_n}(x) - \varrho^* \tau_{f_n}(x) + \int_{\mathbf{X}} h_n(y)\, Q_{f_n}(dy \mid x) \quad \forall x \in \mathbf{X},\ n \in \mathbb{N}_0; \tag{39}$$

(iii) for all $x \in \mathbf{X}$ and $n \in \mathbb{N}_0$,

$$Th_n(x) = C_{f_{n+1}}(x) - \varrho^* \tau_{f_{n+1}}(x) + \int_{\mathbf{X}} h_n(y)\, Q_{f_{n+1}}(dy \mid x) \tag{40}$$

where, for $x \in \mathbf{X}$,

$$Th_n(x) := \min_{a \in A(x)} \Big[C(x,a) - \varrho^* \tau(x,a) + \int_{\mathbf{X}} h_n(y)\, Q(dy \mid x, a) \Big]; \tag{41}$$

(iv) moreover,

$$h_n(x) = h_{n+1}(x) + s_{n+1} \quad \forall x \in N_{n+1} \tag{42}$$

and

$$h_n(x) \geq h_{n+1}(x) + s_{n+1} \quad \forall x \in \mathbf{X}, \tag{43}$$

where $\mu_{f_{n+1}}(N_{n+1}) = 1$ and $s_{n+1} := \mu_{f_{n+1}}(h_n - h_{n+1})$.

Now define

$$N := \bigcap_{n=1}^{\infty} N_n,$$

and observe that $N \neq \emptyset$, since $\varphi(N_n) = \varphi(\mathbf{X}) > 0$ for all $n \in \mathbb{N}$ [where φ is the common irreducibility measure in Assumption 3.2(b)]. Let z be a fixed but arbitrary state in N and define

$$h_n^*(x) := h_n(x) - h_n(z), \quad x \in \mathbf{X},\ n \in \mathbb{N}_0.$$

Then, from (42)–(43), we see that for each $n \in \mathbb{N}_0$,

$$h_n^*(\cdot) \geq h_{n+1}^*(\cdot). \tag{44}$$

Define

$$h_*(x) := \lim_{n \to \infty} h_n^*(x) = \inf_n h_n^*(x), \quad x \in \mathbf{X}. \tag{45}$$

Then, noting that $h_* \in B_W(\mathbf{X})$ [see (32)] and using similar arguments to those in the proof of Lemma 4.4, we have

$$h_*(x) \geq \min_{a\in A(x)}\Big[C(x,a)-\varrho^*\tau(x,a)+\int_{\mathbf{X}} h_*(y)\,Q(dy\,|\,x,a)\Big] \quad \forall x \in \mathbf{X}. \tag{46}$$

Next, we shall prove that the reverse inequality holds. To do this, first observe that (40), (44), (39) and (45) yield, for all x in $\mathbf{X}$,

$$\begin{aligned} Th_n^*(x) &= C_{f_{n+1}}(x) - \varrho^*\tau_{f_{n+1}}(x) + \int_{\mathbf{X}} h_n^*(y)\,Q_{f_{n+1}}(dy\,|\,x) \\ &\geq C_{f_{n+1}}(x) - \varrho^*\tau_{f_{n+1}}(x) + \int_{\mathbf{X}} h_{n+1}^*(y)\,Q_{f_{n+1}}(dy\,|\,x) \\ &= h_{n+1}^*(x) \geq h_*(x). \end{aligned}$$

In consequence, for all $(x,a) \in \mathbf{K}$ and $n \in \mathbb{N}_0$,

$$C(x,a) - \varrho^*\tau(x,a) + \int_{\mathbf{X}} h_n^*(y)\,Q(dy\,|\,x,a) \geq h_*(x).$$

Thus, if we let n go to infinity, the Dominated Convergence Theorem implies

$$C(x,a) - \varrho^*\tau(x,a) + \int_{\mathbf{X}} h_*(y)\,Q(dy\,|\,x,a) \geq h_*(x),$$

from which we have

$$\min_{a\in A(x)}\Big[C(x,a) - \varrho^*\tau(x,a) + \int_{\mathbf{X}} h_*(y)\,Q(dy\,|\,x,a)\Big] \geq h_*(x) \quad \forall x \in \mathbf{X}.$$

Therefore, combining this with (46), we see that $h_*(\cdot)$ satisfies the Average Cost Optimality Equation

$$h_*(x) = \min_{a\in A(x)}\Big[C(x,a) - \varrho^*\tau(x,a) + \int_{\mathbf{X}} h_*(y)\,Q(dy\,|\,x,a)\Big] \quad \forall x \in \mathbf{X}. \tag{47}$$

Finally, the Measurable Selection Theorem guarantees the existence of a policy $f^* \in \mathbf{F}$ such that

$$h_*(x) = C_{f^*}(x) - \varrho^*\tau_{f^*}(x) + \int_{\mathbf{X}} h_*(y)\,Q_{f^*}(dy\,|\,x) \quad \forall x \in \mathbf{X}. \tag{48}$$

Parts (b) and (c) follow from (47) and (48) using standard arguments after noting that

$$E_x^\pi T_n = E_x^\pi \sum_{k=0}^{n-1} \tau(x_k,a_k) \geq n\theta \quad \forall x \in \mathbf{X}, \pi \in \Pi,$$

where θ is the constant in Assumption 3.1(e1). ∎

5. Proof of Theorem 3.7. *Throughout this section we suppose that the assumptions of Theorem* 3.7 *hold.* Consider the function

$$w(x) := W^{1/p}(x), \qquad x \in \mathbf{X}, \tag{49}$$

where p is as in Assumption 3.6, and the normed linear space $B_w(\mathbf{X})$ consists of measurable functions u on $\mathbf{X}$ such that

$$\|u\|_w := \sup_{x\in\mathbf{X}} \frac{|u(x)|}{w(x)} < \infty.$$

Then, using the inequality

$$(a+b)^r \le a^r + b^r \quad \forall a, b \ge 0 \text{ and } 0 \le r \le 1, \tag{50}$$

with $r = 1/p$, and Jensen's inequality, we obtain

$$\int_{\mathbf{X}} w(y)\, Q_f(dy \,|\, x) \le B'_f w(x) + b'_f \mathbf{I}_{K_f}(x) \quad \forall f \in \mathbf{F},\ x \in \mathbf{X}, \tag{51}$$

with $B'_f := B_f^{1/p}$ and $b'_f := b_f^{1/p}$. On the other hand, from Assumption 3.6, we have

$$|C(x,a)| \le r_1^{1/p} w(x) \quad \text{and} \quad \theta^{1/p} < \tau(x,a) \le r_1^{1/p} w(x) \quad \forall (x,a) \in \mathbf{K}. \tag{52}$$

Therefore, as in Remark 3.3 and Lemma 4.2, we have the following:

LEMMA 5.1. *For each stationary policy* $f \in \mathbf{F}$:

(a) $\{x_n\}$ *is* w*-geometrically ergodic*; *that is*, *there exist positive constants* $\widehat{\gamma}_f < 1$ *and* $\widehat{M}_f < \infty$ *such that*

$$\Big|\int_{\mathbf{X}} u(y)\, Q_f^n(dy \,|\, x) - \mu_f(u)\Big| \le \|u\|_w \widehat{M}_f \widehat{\gamma}_f^n w(x) \quad \forall x \in \mathbf{X},\ u \in B_w(\mathbf{X}),\ n \in \mathbb{N};$$

(b) *the function* h_f *in* (30) *belongs to* $B_w(\mathbf{X})$; *hence*, *the solution* $h_*(\cdot)$ *to the Average Cost Optimality Equation* (17) *is in* $B_w(\mathbf{X})$.

The next two lemmas play a key role in the proof of Theorem 3.7. Their proofs are similar to those of Lemmas 4.3 and 4.4 in [17].

LEMMA 5.2. *For each policy* $\pi \in \Pi$ *and initial state* $x \in \mathbf{X}$, *we have*

(a) $E_x^\pi \sum_{k=1}^\infty k^{-p} W(x_k) < \infty$;

hence, *the following statements hold* P_x^π*-a.s.*:

(b) $\sum_{k=1}^\infty k^{-p} W(x_k) < \infty$;
(c) $k^{-p} W(x_k) \to 0$;
(d) $k^{-1} w(x_k) \to 0$.

Proof. Since $W \ge w \ge 1$, it is clear that (a)$\Rightarrow$(b)$\Rightarrow$(c)$\Rightarrow$(d). Thus, it suffices to prove part (a).

To prove (a), let g be a stationary policy as in (23)–(24). Thus,

$$\int_{\mathbf{X}} W(y)\, Q(dy \mid x,a) \le B_g W(x) + b_g \mathbf{I}_{K_g}(x) \qquad \forall (x,a) \in \mathbf{K}.$$

Now, from the properties in Remark 2.2, we see that for each policy π and initial state x,

$$E_x^\pi [W(x_{k+1}) \mid h_k, a_k] = E_x^\pi [W(x_{k+1}) \mid x_k, a_k] \le B_g W(x_k) + b_g, \tag{53}$$

which implies that

$$E_x^\pi W(x_{k+1}) \le B_g E_x^\pi W(x_k) + b_g.$$

Then, by induction, we see that

$$E_x^\pi W(x_k) \le B_g^k W(x) + b_g (1 - B_g)^{-1}. \tag{54}$$

Thus, since $p > 1$,

$$E_x^\pi \sum_{k=1}^{\infty} k^{-p} W(x_k) < \infty \qquad \forall \pi \in \Pi,\ x \in \mathbf{X}. \ \blacksquare$$

Let $\pi \in \Pi$ and $x \in \mathbf{X}$ be arbitrary and define $\mathcal{F}_n := \sigma(h_n, a_n)$, the σ-algebra generated by (h_n, a_n), for each $n \in \mathbb{N}_0$. Moreover, let h_* be a solution to the Average Cost Optimality Equation (17) and define the random variables

$$\begin{aligned} Y_k(\pi,x) &:= h_*(x_k) - E_x^\pi [h_*(x_k) \mid \mathcal{F}_{k-1}] \\ &= h_*(x_k) - \int_{\mathbf{X}} h_*(y)\, Q(dy \mid x_{k-1}, a_{k-1}), \qquad k \in \mathbb{N}, \end{aligned} \tag{55}$$

and

$$M_n(\pi,x) := \sum_{k=1}^{n} Y_k(\pi,x), \qquad n \in \mathbb{N}. \tag{56}$$

Lemma 5.3. *For each $\pi \in \Pi$ and $x \in \mathbf{X}$, the process $(M_n, \mathcal{F}_n)$ is a P_x^π-martingale, and the following statements hold P_x^π-a.s.:*

$$\lim_{n\to\infty} \frac{1}{n} M_n(\pi,x) = 0, \tag{57}$$

$$\lim_{n\to\infty} \frac{1}{\widehat{T}_n} M_n(\pi,x) = 0. \tag{58}$$

Proof. First, note that (57) implies (58). Indeed, from Assumption 3.1(e1), we have

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \tau(x_k, a_k) \ge \theta,$$

which combined with (57), yields (58).

We next prove that (57) holds. To do this, fix $\pi \in \Pi$ and $x \in \mathbf{X}$, and observe that

$$|Y_k(\pi, x)| \le |h_*(x_k)| + E_x^\pi[|h_*(x_k)| \mid x_{k-1}, a_{k-1}], \quad k \in \mathbb{N}.$$

Thus,

$$|Y_k(\pi, x)| \le \|h_*\|_w \{w(x_k) + E_x^\pi[w(x_k) \mid x_{k-1}, a_{k-1}]\}, \quad k \in \mathbb{N}, \tag{59}$$

from which, using (54), we see that

$$E_x^\pi |Y_k(\pi, x)| \le 2\|h_*\|_w E_x^\pi w(x_k) \le 2\|h_*\|_w E_x^\pi W(x_k) < \infty.$$

Hence,

$$E_x^\pi |M_n(\pi, x)| < \infty \quad \text{for every } n \in \mathbb{N}.$$

Now, since $M_n(\pi, x)$ is $\mathcal{F}_n$-measurable, from (55) we conclude that

$$E_x^\pi[M_{n+1}(\pi, x) - M_n(\pi, x) \mid \mathcal{F}_n] = 0 \quad P_x^\pi\text{-a.s.},$$

that is, $(M_n(\pi, x), \mathcal{F}_n)$ is a (P_x^π-) martingale.

Thus, (57) follows from the Strong Law of Large Numbers for martingales ([12], Theorem 2.18, p. 35) provided that

$$\sum_{k=1}^\infty k^{-p} E_x^\pi[|Y_k(\pi, x)|^p \mid \mathcal{F}_{k-1}] < \infty \quad P_x^\pi\text{-a.s.} \tag{60}$$

To prove (60), we use the inequality

$$(a+b)^p \le 2^{p-1}(a^p + b^p) \quad a, b \ge 0,\ p \ge 1,$$

combined with the fact that $w^p(\cdot) = W(\cdot)$ and Jensen's inequality to deduce from (59) that

$$E_x^\pi[|Y_k(\pi, x)|^p \mid \mathcal{F}_{k-1}] \le 2^p \|h_*\|_w^p E_x^\pi[W(x_k) \mid \mathcal{F}_{k-1}] \quad \forall k \in \mathbb{N}.$$

Then, from (53),

$$E_x^\pi[|Y_k(\pi, x)|^p \mid \mathcal{F}_{k-1}] \le 2^p \|h_*\|_w^p (B_g + b_g) W(x_{k-1}),$$

which implies that

$$\begin{aligned}\sum_{k=1}^\infty k^{-p} E_x^\pi[|Y_k(\pi, x)|^p \mid \mathcal{F}_{k-1}] &\le K\Big\{W(x_0) + \sum_{k=2}^\infty k^{-p} W(x_{k-1})\Big\} \\ &\le K\Big\{W(x_0) + \sum_{k=1}^\infty k^{-p} W(x_k)\Big\},\end{aligned}$$

where $K := 2^p \|h_*\|_w^p (B_g + b_g)$. Hence, from Lemma 5.2(b), we see that (60) holds. ■

Proof of Theorem 3.7. (a) Define the "discrepancy" function on $\mathbf{K}$:

$$D(x, a) := C(x, a) - \varrho^* \tau(x, a) + \int_{\mathbf{X}} h_*(y)\, Q(dy \mid x, a) - h_*(x),$$

and note that it is nonnegative because $h_*(\cdot)$ is a solution to the Average Cost Optimality Equation (17). Then observe that for every $n \in \mathbb{N}$,

$$\sum_{k=0}^{n-1} D(x_k, a_k) = \sum_{k=0}^{n-1} C(x_k, a_k) - \varrho^* \widehat{T}_n - M_n(\pi, x) - h_*(x_0) + h_*(x_n),$$

which, since D is nonnegative, yields

$$\frac{1}{\widehat{T}_n} \sum_{k=0}^{n-1} C(x_k, a_k) \geq \varrho^* + \frac{1}{\widehat{T}_n} [M_n(\pi, x) + h_*(x_0) - h_*(x_n)].$$

Hence, from Lemma 5.2(d) and (58), we get

$$J_1(\pi, x) \geq \liminf_{n\to\infty} \frac{1}{\widehat{T}_n} \sum_{k=0}^{n-1} C(x_k, a_k) \geq \varrho^* \quad P_x^\pi\text{-a.s.} \tag{61}$$

(b) This part is an immediate consequence of part (a), Remark 4.1(c) and Theorem 3.5. ∎

6. Proof of Theorem 3.8. *In this section we suppose that the assumptions of Theorem* 3.8 *hold.* For each $\pi \in \Pi$ and $x \in \mathbf{X}$, define the random variables

$$Z_k(\pi, x) := \delta_k - E_x^\pi[\delta_k \mid \mathcal{F}_{k-1}], \quad k \in \mathbb{N}, \tag{62}$$

where $\mathcal{F}_n = \sigma(h_n, a_n)$ for each $n \in \mathbb{N}$, and

$$M_n'(\pi, x) := \sum_{k=1}^{n} Z_k(\pi, x), \quad n \in \mathbb{N}. \tag{63}$$

Observe, from Remark 2.2, that

$$Z_k(\pi, x) = \delta_k - \tau(x_{k-1}, a_{k-1}), \quad M_n'(\pi, x) = T_n - \widehat{T}_n, \quad k, n \in \mathbb{N}. \tag{64}$$

Theorem 3.8 is an immediate consequence of Lemmas 6.1 and 6.2.

Lemma 6.1. *For each $\pi \in \Pi$ and $x \in \mathbf{X}$, the process $(M_n'(\pi, x), \mathcal{F}_n)$ is a $(P_x^\pi$-$)$ martingale and*

$$\lim_{n\to\infty} \frac{1}{n} M_n'(\pi, x) = 0 \quad P_x^\pi\text{-a.s.} \tag{65}$$

Proof. Let $\pi \in \Pi$ and $x \in \mathbf{X}$ be fixed but arbitrary. To prove that $(M_n'(\pi, x), \mathcal{F}_n)$ is a $(P_x^\pi$-$)$ martingale, from (62)–(63), it only remains to check that

$$E_x^\pi |M_n'(\pi, x)| < \infty \quad \forall n \in \mathbb{N}_0.$$

This follows on noting that

$$|Z_k(\pi, x)| \leq \delta_k + \tau(x_{k-1}, a_{k-1}) \quad \forall n \in \mathbb{N}.$$

Thus,

$$E_x^\pi[|Z_k(\pi,x)|\,|\,\mathcal{F}_{k-1}] \le 2\tau(x_{k-1},a_{k-1}) \le 2r_2 W(x_{k-1}),$$

where r_2 is the constant in (22). Then, by (54),

$$E_x^\pi|Z_k(\pi,x)| \le 2r_2 E_x^\pi W(x_{k-1}) < \infty.$$

Hence, $(M'_n(\pi,x),\mathcal{F}_n)$ is a (P_x^π-) martingale.

As in the proof of Lemma 5.3, (65) follows from the Strong Law of Large Numbers for martingales ([12], Theorem 2.18, p. 35) provided that

$$\sum_{k=1}^{\infty} k^{-p} E_x^\pi[|Z_k(\pi,x)|^p\,|\,\mathcal{F}_{k-1}] < \infty \quad P_x^\pi\text{-a.s.} \tag{66}$$

We next prove (66). In order to do this, observe

$$|Z_k(\pi,x)|^p \le [\delta_k + \tau(x_{k-1},a_{k-1})]^p \le 2^{p-1}[\delta_k^p + \tau^p(x_{k-1},a_{k-1})].$$

Then

$$E_x^\pi[|Z_k(\pi,x)|^p\,|\,\mathcal{F}_{k-1}] \le 2^{p-1}[\eta(x_{k-1},a_{k-1}) + \tau^p(x_{k-1},a_{k-1})].$$

Now, from (22) and Jensen's inequality, we have

$$E_x^\pi[|Z_k(\pi,x)|^p\,|\,\mathcal{F}_{k-1}] \le 2^p\eta(x_{k-1},a_{k-1}) \le 2^p r_2 W(x_{k-1}).$$

Thus, proceeding as in the proof of Lemma 5.3, we conclude that

$$\sum_{k=1}^{\infty} k^{-p} E_x^\pi[|Z_k(\pi,x)|^p\,|\,\mathcal{F}_{k-1}] \le \widehat{L}\Big\{W(x_0) + \sum_{k=1}^{\infty} k^{-p} W(x_k)\Big\} < \infty$$

holds P_x^π-a.s., where $\widehat{L} := 2^p r_2(B_g + b_g)$. ∎

LEMMA 6.2. *For each policy $\pi \in \Pi$ and $x \in \mathbf{X}$, the following facts hold P_x^π-a.s.:*

(a) $\lim_{n\to\infty} T_n = \infty$;
(b) $\lim_{n\to\infty} T_n/\widehat{T}_n = 1$;
(c) $\lim_{n\to\infty} \widehat{T}_{n+1}/\widehat{T}_n = \lim_{n\to\infty} T_{n+1}/T_n = 1$;
(d) $\lim_{t\to\infty} T_{\eta(t)+1}/T_{\eta(t)} = \lim_{t\to\infty}(1/t)T_{\eta(t)} = 1$.

Proof. (a) Let $\pi \in \Pi$ and $x \in \mathbf{X}$ be fixed but arbitrary. We see, from (65) and Assumption 3.1(e1), that

$$\liminf_{n\to\infty} \frac{1}{n} T_n = \liminf_{n\to\infty} \frac{1}{n}\widehat{T}_n \ge \theta \quad P_x^\pi\text{-a.s.,} \tag{67}$$

which implies that

$$\lim_{n\to\infty} T_n = \infty.$$

(b) Using (65) again and (67), we have

$$\lim_{n\to\infty} \frac{M'_n}{T_n} = \lim_{n\to\infty}\left[1 - \frac{\widehat{T}_n}{T_n}\right] = 0 \quad P_x^\pi\text{-a.s.,}$$

or equivalently,

$$\lim_{n\to\infty} \frac{\widehat{T}_n}{T_n} = 1 \qquad P_x^\pi\text{-a.s.}$$

(c) Observe that, from (49) and (52), the following inequalities hold:

$$1 \le \frac{\widehat{T}_{n+1}}{\widehat{T}_n} = 1 + \frac{\tau(x_n, a_n)}{\widehat{T}_n} \le 1 + \frac{1}{\theta^{1/p}} \frac{1}{n} \tau(x_n, a_n) \le 1 + \frac{r_1}{\theta^{1/p}} \frac{1}{n} w(x_n).$$

Thus, Lemma 5.2(d) yields

$$\lim_{n\to\infty} \frac{\widehat{T}_{n+1}}{\widehat{T}_n} = 1 \qquad P_x^\pi\text{-a.s.} \tag{68}$$

Now observe that

$$\frac{T_{n+1}}{T_n} = \frac{T_{n+1}}{\widehat{T}_{n+1}} \frac{\widehat{T}_{n+1}}{\widehat{T}_n} \frac{\widehat{T}_n}{T_n}; \tag{69}$$

hence, from part (b), (68) and (69),

$$\lim_{n\to\infty} \frac{T_{n+1}}{T_n} = 1 \qquad P_x^\pi\text{-a.s.}$$

(d) Since $T_n < \infty$, for every $n \in \mathbb{N}_0$, and $T_n \to \infty$ P_x^π-a.s., we deduce that for every nonnegative real $t \in \mathbb{R}^+$ there exists a nonnegative integer $k \in \mathbb{N}_0$, depending on the realization of the process, such that

$$T_k \le t < T_{k+1}.$$

Thus,

$$\frac{T_{\eta(t)+1}}{T_{\eta(t)}} = \frac{T_{k+1}}{T_k} \qquad \text{if } T_k \le t < T_{k+1}.$$

Hence, from (c),

$$\lim_{t\to\infty} \frac{T_{\eta(t)+1}}{T_{\eta(t)}} = 1 \qquad P_x^\pi\text{-a.s.} \tag{70}$$

To prove the second equality, observe that

$$T_{\eta(t)} \le t < T_{\eta(t)+1} \qquad \forall t > 0;$$

equivalently,

$$1 \le \frac{t}{T_{\eta(t)}} < \frac{T_{\eta(t)+1}}{T_{\eta(t)}}.$$

Thus, from (70), we see that

$$\lim_{t\to\infty} \frac{T_{\eta(t)}}{t} = 1 \qquad P_x^\pi\text{-a.s.} \ \blacksquare$$

Proof of Theorem 3.8. (a) Let $\pi \in \Pi$ and $x \in \mathbf{X}$ be fixed but arbitrary. From Lemma 6.2(b), the following holds P_x^π-a.s.:

$$\begin{aligned} J_0(\pi, x) &= \limsup_{n\to\infty} \left\{ \frac{\widehat{T}_n}{T_n} \frac{1}{\widehat{T}_n} \sum_{k=0}^{n-1} C(x_k, a_k) \right\} \\ &= \limsup_{n\to\infty} \frac{1}{\widehat{T}_n} \sum_{k=0}^{n-1} C(x_k, a_k) = J_1(\pi, x), \end{aligned}$$

which proves the first equality. To prove the second one, first note that Lemma 6.2(c) yields

$$\begin{aligned} J_2(\pi, x) &= \limsup_{t\to\infty} \frac{1}{t} \sum_{k=0}^{\eta(t)} C(x_k, a_k) \\ &= \limsup_{t\to\infty} \left\{ \frac{T_{\eta(t)+1}}{T_{\eta(t)}} \frac{T_{\eta(t)}}{t} \frac{1}{T_{\eta(t)+1}} \sum_{k=0}^{\eta(t)} C(x_k, a_k) \right\} \\ &= \limsup_{t\to\infty} \frac{1}{T_{\eta(t)+1}} \sum_{k=0}^{\eta(t)} C(x_k, a_k). \end{aligned}$$

Thus, using similar arguments to those in the proof of Lemma 6.2(d), we see that

$$J_0(\pi, x) = \limsup_{n\to\infty} \frac{1}{T_n} \sum_{k=0}^{n-1} C(x_k, a_k) = \limsup_{t\to\infty} \frac{1}{T_{\eta(t)+1}} \sum_{k=0}^{\eta(t)} C(x_k, a_k),$$

which proves that $J_0(\pi, x) = J_2(\pi, x)$ P_x^π-a.s.

(b) This part is a direct consequence of part (a) and Theorem 3.7(b). ∎

References

[1] R. B. Ash, *Real Analysis and Probability*, Academic Press, New York, 1972.
[2] S. Bhatnagar and V. S. Borkar, *A convex analytic framework for ergodic control of semi-Markov processes*, Math. Oper. Res. 20 (1995), 923–936.
[3] R. N. Bhattacharya and M. Majumdar, *Controlled semi-Markov models under long-run average rewards*, J. Statist. Plann. Inference 22 (1989), 223–242.
[4] B. S. Borkar, *Topics in Controlled Markov Chains*, Pitman Res. Notes Math. Ser. 240, Longman Sci. Tech., 1991.
[5] R. Cavazos-Cadena and E. Fernández-Gaucherand, *Denumerable controlled Markov chains with average reward criterion*: *Sample path optimality*, Z. Oper. Res. Math. Methods Oper. Res. 41 (1995), 89–108.
[6] A. Federgruen, A. Hordijk and H. C. Tijms, *Denumerable state semi-Markov decision processes with unbounded costs*, *average cost criterion*, Stochastic Process. Appl. 9 (1979), 223–235.

[7] A. Federgruen, P. J. Schweitzer and H. C. Tijms, *Denumerable undiscounted semi-Markov decision processes with unbounded rewards*, Math. Oper. Res. 8 (1983), 298–213.

[8] A. Federgruen and H. C. Tijms, *The optimality equation in average cost denumerable state semi-Markov decision problems. Recurrence conditions and algorithms*, J. Appl. Probab. 15 (1978), 356–373.

[9] E. A. Feinberg, *Constrained semi-Markov decision processes with average rewards*, Z. Oper. Res. (Math. Methods Oper. Res.) 39 (1994), 257–288.

[10] P. W. Glynn and S. P. Meyn, *A Liapunov bound for solutions of Poisson's equation*, Ann. Probab. 24 (1996), 916–931.

[11] 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.

[12] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application*, Academic Press, 1980.

[13] U. G. Haussman, *On the optimal long-run control of Markov renewal processes*, J. Math. Anal. Appl. 36 (1971), 123–140.

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

[15] —, *Further criteria for positive Harris recurrence of Markov chains*, Proc. Amer. Math. Soc., to appear.

[16] 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.

[17] 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., to appear.

[18] M. Kurano, *Semi-Markov decision processes and their applications in replacement models*, J. Oper. Res. Soc. Japan 28 (1985), 18–29.

[19] —, *Average optimal adaptive policies in semi-Markov decision processes including an unknown parameter*, ibid., 252–266.

[20] J. B. Lasserre, *Sample-path average optimality for Markov control processes*, IEEE Trans. Automat. Control, to appear.

[21] S. A. Lippman, *Semi-Markov decision processes with unbounded rewards*, Management Sci. 19 (1973), 717–731.

[22] —, *On dynamic programming with unbounded rewards*, ibid. 21 (1975), 1225–1233.

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

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

[25] M. L. Puterman, *Markov Decision Processes*, Wiley, New York, 1994.

[26] S. M. Ross, *Average cost semi-Markov decision processes*, J. Appl. Probab. 7 (1979), 649–656.

[27] M. Schäl, *On the second optimality equation for semi-Markov decision models*, Math. Oper. Res. 17 (1992), 470–486.

[28] P. J. Schweitzer, *Iterative solutions of the functional equations of undiscounted Markov renewal programming*, J. Math. Anal. Appl. 34 (1971), 495–501.

[29] L. I. Sennott, *Average cost semi-Markov decision processes and the control of queueing systems*, Probab. Engrg. Inform. Sci. 3 (1989), 247–272.

[30] O. Vega-Amaya, *Sample path average optimality of Markov control processes with strictly unbounded cost*, Appl. Math. (Warsaw) 26 (1999), 363–381.

[31] —, *Markov control processes in Borel spaces: undiscounted cost criteria*, doctoral thesis, UAM-Iztapalapa, México, 1998 (in Spanish).

Oscar Vega-Amaya and Fernando Luque-Vásquez
Departamento de Matemáticas
Universidad de Sonora
Blvd. Transversal y Rosales s/n
83000 Hermosillo, Sonora, México
E-mail: ovega@gauss.mat.uson.mx
fluque@gauss.mat.uson.mx

Received on 2.11.1999;
revised version on 28.2.2000