

F. LUQUE-VÁSQUEZ (Sonora)
O. HERNÁNDEZ-LERMA (México)

SEMI-MARKOV CONTROL MODELS WITH AVERAGE COSTS

Abstract. This paper studies semi-Markov control models with Borel state and control spaces, and unbounded cost functions, under the average cost criterion. Conditions are given for (i) the existence of a solution to the average cost optimality equation, and for (ii) the existence of strong optimal control policies. These conditions are illustrated with a semi-Markov replacement model.

1. Introduction. This paper deals with semi-Markov control models (SMCMs) with Borel state and action spaces, possibly unbounded costs, and the average cost (AC) criterion. Under suitable assumptions on the transition law, the cost functions and the distribution of the sojourn times, our main result (Theorem 3.5) shows the existence of a solution to the average cost optimality equation and the existence of a strong AC-optimal stationary policy.

Our approach is a combination of facts and techniques from different sources. For instance, our Assumptions 3.1 and 3.2 are partly based on hypotheses previously used in [7] for Markov control models (MCMs), but with very important differences:

(a) In contrast to [7, Assumption 2.3], we do not impose the condition of positive Harris recurrence. Instead, in Lemma 4.1 below we prove that the latter condition is a consequence of our Assumptions 3.2 and 3.3. (The former is the same as [7, Assumption 2.4].)

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

Key words and phrases: semi-Markov control models, average cost, policy iteration (or Howard's algorithm), replacement models.

This research was partially supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT) Grants 3871P-E9607 and 3115P-E9608.

(b) We consider cost-per-stage functions that may take positive as well negative values. Thus, since these functions may also be unbounded, we cannot use the standard vanishing-discount argument (as in the proof of Theorem 2.6 in [7]) to obtain a stationary policy, say f^* , which is AC-optimal in the class of all policies; we deduce instead that f^* is AC-optimal in the (much smaller) subclass of stationary policies.

(c) We do not require a Lipschitz-like condition as in [7, Assumption 2.7], and, therefore, we cannot use the Ascoli Theorem to obtain a solution to the AC-optimality equation (ACOE); see Theorem 2.8 and §6 in [7]. In the present paper, to obtain the ACOE (in (6) below) we use a version of the policy iteration (or Howard's) algorithm which at the initialization step uses the policy f^* mentioned in (b). In this part we essentially follow [24], [10, §10.3.B].

Another key step in the proof of our main result, Theorem 3.5, is that we use a data transformation to associate a MCM with our original SMCM. This transformation has been used by several authors (for example, [4, 5, 21]) to study SMCs with a countable state space. Finally, as is well known, for SMCs the AC criterion itself can be defined in several different forms. Here we use the form in Definition 2.3, which has been previously studied by many authors, for example [2–5, 8, 11, 16, 18, 19, 21–23]. Note, however, that most of these references deal with a countable state space (see also [12, 20]), whereas those that consider general Borel spaces follow the so-called vanishing-discount approach [2, 13, 19] and/or use bounded costs [11].

The remainder of the paper is organized in six sections. In Section 2 we introduce the SMCM we will be dealing with, together with the performance criterion. In Section 3 we introduce the hypotheses and present our main result, Theorem 3.5. Section 4 presents some preliminary facts, and in Section 5 we introduce the data transformation and some results related to the associated MCM. The proof of Theorem 3.5 is given in Section 6. The paper concludes in Section 7 with an example of a replacement model that satisfies the assumptions of our main result.

Terminology and notation. 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 fixed $y \in Y$ 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 real) numbers.

2. The semi-Markov control model. The *semi-Markov control model* $(X, A, \{A(x) : x \in X\}, Q, F, D, d)$ consists of the *state space* X and the

control set A , which are assumed to be Borel spaces. For each $x \in X$, there exists a non-empty Borel subset $A(x)$ of A whose elements are the *admissible controls* if the state of the system is x . It is assumed that the set

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

is a Borel subset of $X \times A$ and that it contains the graph of a measurable function $f : X \rightarrow A$. Moreover, Q is a stochastic kernel on X given \mathbf{K} and represents the *transition law*, whereas for each $(x, a) \in \mathbf{K}$, $F(\cdot | x, a)$ is a distribution function and represents the *distribution of the sojourn times*. The measurable functions D and d on \mathbf{K} are cost functions.

If x_n and a_n denote the state of the system and the control action at the n th decision time T_n ($n = 0, 1, \dots$, with $T_0 := 0$), then $D(x_n, a_n)$ stands for the immediate cost incurred at time T_n , and $d(x_n, a_n)$ is the cost rate during the interval $[T_n, T_{n+1})$. The corresponding transition (or sojourn) time $\delta_{n+1} := T_{n+1} - T_n$ ($n = 0, 1, \dots$) has the distribution $F(\cdot | x_n, a_n)$.

DEFINITION 2.1. \mathbf{F} denotes the set of measurable functions $f : X \rightarrow A$ such that $f(x) \in A(x)$ for $x \in X$.

For each $n = 0, 1, \dots$ define the space of *admissible histories* until the n th transition by $H_0 := X$ and

$$H_n := (\mathbf{K} \times \mathbb{R})^n \times X \quad \text{for } n = 1, 2, \dots$$

DEFINITION 2.2. A *control policy* (or briefly, a *policy*) is a sequence $\pi = \{\pi_n\}$ of stochastic kernels π_n on A given H_n satisfying the constraint $\pi_n(A(x_n) | i_n) = 1$ for all $i_n = (x_0, a_0, \delta_1, \dots, x_{n-1}, a_{n-1}, \delta_n, x_n) \in H_n$, $n \in \mathbb{N}_0$. The collection of all policies is denoted by Π . A policy $\pi = \{\pi_n\}$ is said to be a *stationary policy* if there exists $f \in \mathbf{F}$ such that $\pi_n(\cdot | i_n)$ is concentrated at $f(x_n)$ for each n . We will identify \mathbf{F} with the class of stationary policies.

Let (Ω, \mathcal{F}) be the measurable space consisting of the sample space $\Omega := (X \times A \times \mathbb{R}_+)^{\infty}$ and the corresponding product σ -algebra \mathcal{F} . By a theorem of C. Ionescu Tulcea (see [1], Theorem 2.7.2, p. 109), for each initial state $x \in X$ and each policy $\pi \in \Pi$, there exists a probability measure P_x^{π} such that for all $B \in \mathcal{B}(A)$, $C \in \mathcal{B}(X)$ and $i_n = (x_0, a_0, \delta_1, \dots, x_{n-1}, a_{n-1}, \delta_n, x_n)$ in H_n , with $n = 0, 1, \dots$, we have

$$\begin{aligned} P_x^{\pi}(x_0 = x) &= 1, \\ P_x^{\pi}(a_n \in B | i_n) &= \pi_n(B | i_n), \\ (1) \quad P_x^{\pi}(x_{n+1} \in C | i_n, a_n, \delta_{n+1}) &= Q(C | x_n, a_n), \end{aligned}$$

and

$$P_x^{\pi}(\delta_{n+1} \leq t | i_n, a_n) = F(t | x_n, a_n).$$

E_x^{π} denotes the expectation operator with respect to P_x^{π} .

For a policy $\pi \in \Pi$, the random variable x_n describes the state of the system at the time of the n th transition, when the controls are chosen according to the policy π . It is clear that, in general, such a state depends on the evolution of the system in the first $n - 1$ transitions; however, for a stationary policy f , $\{x_n\}$ is a Markov chain with transition probability $Q(\cdot | x, f(x))$. This is a consequence of the properties of the conditional expectation and the Markov-like property (1).

The *mean holding time* in the state x when the control $a \in A(x)$ is chosen is given by

$$\tau(x, a) := \int_{\mathbb{R}} t F(dt | x, a).$$

DEFINITION 2.3. For $x \in X$ and $\pi \in \Pi$, define the *expected average cost* by

$$J(\pi, x) := \limsup_{n \rightarrow \infty} \frac{E_x^\pi \{ \sum_{k=0}^{n-1} [D(x_k, a_k) + \delta_{k+1} d(x_k, a_k)] \}}{E_x^\pi(T_n)}.$$

The function

$$J(x) := \inf_{\pi \in \Pi} J(\pi, x)$$

is the *optimal average cost function*, and a policy $\pi^* \in \Pi$ is said to be *average cost optimal* (AC-optimal) if $J(x) = J(\pi^*, x)$ for all $x \in X$.

Using properties of the conditional expectation we can write

$$(2) \quad J(\pi, x) = \limsup_{n \rightarrow \infty} \frac{E_x^\pi [\sum_{k=0}^{n-1} C(x_k, a_k)]}{E_x^\pi [\sum_{k=0}^{n-1} \tau(x_k, a_k)]},$$

where $C(x, a) := D(x, a) + \tau(x, a)d(x, a)$.

REMARK 2.4. A *Markov control model* (MCM) is the particular case of a SMCM in which $\delta_n \equiv 1$ for all n , so that $\tau(x, a) = 1$ and

$$J(\pi, x) = \limsup_{n \rightarrow \infty} \frac{E_x^\pi [\sum_{k=0}^{n-1} C(x_k, a_k)]}{n}.$$

REMARK 2.5. For a stationary policy $f \in \mathbf{F}$, we will use the abbreviated notation (x, f) in lieu of $(x, f(x))$. In particular, $C(x, f) := C(x, f(x))$, and $Q(\cdot | x, f) := Q(\cdot | x, f(x))$.

3. The optimality conditions and main result. The problem we are concerned with is to show the existence of AC-optimal policies, which requires imposing suitable assumptions on the model. As was already mentioned in §1, the assumptions we use here are partly based on the hypotheses introduced in [7] for MCMs.

ASSUMPTION 3.1. (a) $A(x)$ is compact for each state $x \in X$.

(b) $C(x, \cdot)$ is l.s.c. (lower semicontinuous) on $A(x)$ for each $x \in X$; moreover, there exist a measurable function $v : X \rightarrow [1, \infty)$ and $\gamma \in \mathbb{R}$ such that for each $x \in X$,

$$(3) \quad \sup_{a \in A(x)} |C(x, a)| \leq \gamma v(x),$$

and

$$a \mapsto \int v(y) Q(dy | x, a) \text{ is continuous on } A(x).$$

(c) For each $x \in X$, the transition law Q is strongly continuous on $A(x)$; that is, for each bounded measurable function $u : X \rightarrow \mathbb{R}$, the function

$$a \mapsto \int u(y) Q(dy | x, a)$$

is continuous on $A(x)$.

(d) For each $x \in X$, $\tau(x, \cdot)$ is continuous on $A(x)$, and there exist positive constants m and M such that

$$(4) \quad m \leq \tau(x, a) \leq M \quad \forall (x, a) \in \mathbf{K}.$$

ASSUMPTION 3.2. There exist a probability measure ν on X and a number $0 < \alpha < 1$ for which the following holds: For each $f \in \mathbf{F}$ there is a nonnegative measurable function ϕ_f on X such that for every $x \in X$ and $B \in \mathcal{B}(X)$:

$$(a) \quad Q(B | x, f) \geq \phi_f(x) \nu(B).$$

(b) $\int v(y) Q(dy | x, f) \leq \phi_f(x) \|\nu\|_v + \alpha v(x)$, where the norm is defined by $\|\nu\|_v := \int v(y) \nu(dy) < \infty$.

$$(c) \quad \inf_{\mathbf{F}} \int \phi_f(y) \nu(dy) > 0.$$

Note that taking $B = X$ in Assumption 3.2(a), we obtain $\phi_f(x) \leq 1$ for all $x \in X$.

ASSUMPTION 3.3. There is a σ -finite measure ψ on X with respect to which, for each $f \in \mathbf{F}$, the Markov chain defined by $Q(\cdot | \cdot, f)$ is ψ -irreducible.

We denote by L_v^∞ the normed linear space of measurable functions $u : X \rightarrow \mathbb{R}$ with norm

$$\|u\|_v := \sup_{x \in X} \frac{|u(x)|}{v(x)} < \infty.$$

DEFINITION 3.4. A pair $(j, h(\cdot))$ consisting of a real number j and a measurable function h is said to be a *solution to the average cost optimality equation* (ACOE) if, for all $x \in X$,

$$(5) \quad h(x) = \min_{a \in A(x)} \left\{ C(x, a) - j\tau(x, a) + \int h(y) Q(dy | x, a) \right\}.$$

In the case of a MCM (as in Remark 2.4), $\tau(x, a) \equiv 1$ in (5).

We are now ready to state our main result, Theorem 3.5. The proof is presented in Section 6, and Section 7 shows an example for which the hypotheses of the theorem are satisfied.

THEOREM 3.5. *If Assumptions 3.1–3.3 hold, then there exist $j \in \mathbb{R}$, $h \in L_v^\infty$ and $f \in \mathbf{F}$ such that:*

(a) *$(j, h(\cdot))$ is a solution to the ACOE (5) and, furthermore,*

$$(6) \quad h(x) = C(x, f) - j\tau(x, f) + \int h(y) Q(dy | x, f).$$

(b) *f is AC-optimal and $j = J(x) = J(f, x)$ for all $x \in X$; in fact, f is strong AC-optimal, in the sense that*

$$J(f, x) \leq \liminf_{n \rightarrow \infty} \frac{E_x^\pi [\sum_{k=0}^{n-1} C(x_k, a_k)]}{E_x^\pi(T_n)} \leq J(\pi, x)$$

for each $\pi \in \Pi$ and $x \in X$.

Assumptions 3.1–3.3 are supposed to hold throughout the sequel. In the next two sections we present preliminary results needed to prove Theorem 3.5.

4. Preliminary results. Concerning Lemma 4.1 see the paragraph (a) in Section 1. In the proof of the lemma we use the notions of *small* and *petite* sets, for which the reader is referred to [6, 14, 15], for instance. Also recall that the function ϕ_f in Assumption 3.2 satisfies $0 \leq \phi_f \leq 1$.

LEMMA 4.1. *If Assumptions 3.2 and 3.3 hold, then, for each $f \in \mathbf{F}$, the Markov chain defined by $Q(\cdot | \cdot, f)$ is positive Harris recurrent; hence, in particular, $Q(\cdot | \cdot, f)$ admits a unique invariant probability measure which will be denoted by q_f .*

Proof. Choose an arbitrary stationary policy f . Then, by Assumption 3.2(c), there is a positive number $d = d_f$ for which the set $C = C_f$ given by

$$C := \{x \in X : \phi_f(x) \geq d\}$$

has positive ν -measure. Moreover, by Assumption 3.2(a),

$$Q(B|x, f) \geq d\nu(B) \quad \forall x \in C, B \in \mathcal{B}(X),$$

which implies that C is a small set, and, therefore, petite (by [14, Proposition 5.5.3]). Now note that we may take $d = d_f$ to satisfy $d \leq (1-\alpha)/(2\|\nu\|_v)$. Then, from Assumption 3.2(b) and the fact that $0 \leq \phi_f \leq 1$, we obtain

$$\int v(y) Q(dy | x, f) \leq \alpha v(x) + \|\nu\|_v I_C(x) + d\|\nu\|_v,$$

(where I_C denotes the indicator of C), which yields

$$\int v(y) Q(dy | x, f) \leq \beta v(x) + bI_C(x)$$

with $\beta := (\alpha + 1)/2 < 1$ and $b := \|\nu\|_v$. This inequality and Assumption 3.3 imply that $Q(\cdot | \cdot, f)$ is positive Harris recurrent, by Theorem 2.2 in [6]. ■

From Lemma 4.1 above and Lemma 3.4 in [7] we obtain the following.

LEMMA 4.2. *If Assumptions 3.2 and 3.3 hold, then there exist constants $k \geq 0$ and $0 < \eta < 1$ such that for every $g \in L_v^\infty$ and $x \in X$,*

$$(7) \quad \sup_{f \in \mathbf{F}} \left| \int g(y) Q^n(dy | x, f) - \int g(y) q_f(dy) \right| \leq \|g\|_v k v(x) \eta^n,$$

where $Q^n(\cdot | \cdot, f)$ denotes the n -step transition kernel.

A consequence of Lemma 4.2 is that, for each $f \in \mathbf{F}$, the expected average cost in (2) is a constant:

$$(8) \quad J(f, x) = \frac{\int C(y, f) q_f(dy)}{\int \tau(y, f) q_f(dy)} =: J_f \quad \forall x \in X.$$

On the other hand, for each $n \in \mathbb{N}$ and $f \in \mathbf{F}$, Assumption 3.1(b) gives

$$\left| \int C(y, f) Q^n(dy | x, f) \right| \leq \gamma \int v(y) Q^n(dy | x, f),$$

and, by Assumption 3.2(b),

$$\int v(y) Q^n(dy | x, f) \leq \|\nu\|_v (1 + \alpha + \dots + \alpha^{n-1}) + \alpha^n v(x)$$

since $\phi_f(\cdot) \leq 1$. Hence

$$\sup_{f \in \mathbf{F}} \int C(y, f) q_f(dy) \leq \frac{\gamma \|\nu\|_v}{1 - \alpha} =: \gamma',$$

which in turn yields

$$(9) \quad \sup_{f \in \mathbf{F}} |J_f| \leq \frac{\gamma'}{m},$$

where m is the constant in (4).

LEMMA 4.3. *Suppose that Assumptions 3.1–3.3 hold. Then:*

(a) *For each $f \in \mathbf{F}$ there exists $h_f \in L_v^\infty$ such that*

$$(10) \quad h_f(x) = C(x, f) - J_f \tau(x, f) + \int h_f(y) Q(dy | x, f) \quad \forall x \in X.$$

(b) *If the pair (J'_f, h'_f) , with $J'_f \in \mathbb{R}$ and $h'_f \in L_v^\infty$, is another solution to the equation (10), i.e.,*

$$(11) \quad h'_f(x) = C(x, f) - J'_f \tau(x, f) + \int h'_f(y) Q(dy | x, f) \quad \forall x \in X,$$

then $J'_f = J_f$ and $h'_f(x) = h_f(x) + k$ for all $x \in X$ and some constant k .

(c) *For $x \in X$, $\pi \in \Pi$, and $u \in L_v^\infty$,*

$$(12) \quad \lim_{n \rightarrow \infty} \frac{E_x^\pi u(x_n)}{n} = 0.$$

Proof. (a) Define

$$h_f(x) := \sum_{n=0}^{\infty} E_x^f [C(x_n, f) - J_f \tau(x_n, f)].$$

First we prove that $h_f \in L_v^\infty$. To do this, observe that, using (8) and (7), for each $n = 0, 1, \dots$ we obtain

$$\begin{aligned} & |E_x^f [C(x_n, f) - J_f \tau(x_n, f)]| \\ &= \left| \int C(y, f) Q^n(dy | x, f) - J_f \int \tau(y, f) Q^n(dy | x, f) \right| \\ &\leq \left| \int C(y, f) Q^n(dy | x, f) - \int C(y, f) q_f(dy) \right| \\ &\quad + |J_f| \left| \int \tau(y, f) Q^n(dy | x, f) - \int \tau(y, f) q_f(dy) \right| \\ &\leq Mv(x)\eta^n \end{aligned}$$

for some constant M . Therefore

$$(13) \quad |h_f(x)| \leq \frac{M}{1-\eta} v(x) \quad \forall x \in X,$$

and so $h_f \in L_v^\infty$. Finally, (10) follows from the definition of h_f and the Markov property.

(b) Integrating both sides of (11) with respect to the invariant probability measure q_f , we obtain $J'_f = J_f$. On the other hand, subtracting (11) from (10), it follows that the function $u \in L_v^\infty$ defined by $u(x) := h_f(x) - h'_f(x)$ is an f -harmonic function, i.e.,

$$u(x) = \int u(y) Q(dy | x, f) \quad \text{for all } x \in X.$$

This implies that

$$u(x) = \int u(y) Q^n(dy | x, f) \quad \text{for all } x \in X \text{ and } n = 0, 1, \dots,$$

and, therefore, letting $n \rightarrow \infty$, we deduce from (7) that $u(\cdot)$ is the constant $k := \int u dq_f$.

(c) Using Example 2.6 in [17], it can be shown that for each $x \in X$ and $a \in A(x)$, there exists $f \in \mathbf{F}$ such that $a = f(x)$. Thus, Assumption 3.2(b) implies that

$$\sup_{a \in A(x)} \int v(y) Q(dy | x, a) \leq \|\nu\|_v + \alpha v(x),$$

because $\phi_f(\cdot) \leq 1$. Then, for $u \in L_v^\infty$, $x \in X$ and $\pi \in \Pi$,

$$(14) \quad |E_x^\pi u(x_n)| \leq \|u\|_v \left[\frac{\|\nu\|_v}{1-\alpha} + \alpha^n v(x) \right] \quad \forall n \in \mathbb{N},$$

which yields (12). ■

5. The associated Markov model. The proof of Theorem 3.5 uses, in particular, the transformation introduced in [21] and applied by several authors for SMCs with denumerable state space (see e.g. [4, 5, 21]). In this section we define the transformation in our present context and prove some results related to the Markov control model (MCM) associated with the SMC.

DEFINITION 5.1. Let τ be a real number such that $0 < \tau < m$ (see (4)). Define the function $\widehat{C} : \mathbf{K} \rightarrow \mathbb{R}$ and the stochastic kernel \widehat{Q} on X given \mathbf{K} by

$$(15) \quad \widehat{C}(x, a) := \frac{C(x, a)}{\tau(x, a)}$$

and

$$(16) \quad \widehat{Q}(B | x, a) := \frac{\tau}{\tau(x, a)} Q(B | x, a) + \left(1 - \frac{\tau}{\tau(x, a)}\right) \delta_x(B)$$

where $\delta_x(\cdot)$ is the Dirac measure concentrated at x .

We refer to $(X, A, \{A(x) : x \in X\}, \widehat{Q}, \widehat{C})$ as the *Markov control model associated with the SMC*. The following lemmas essentially show that the associated MCM satisfies Assumptions 3.1–3.3.

LEMMA 5.2. *Parts (b) and (c) of Assumption 3.1 hold when C and Q are replaced by \widehat{C} and \widehat{Q} , respectively.*

PROOF. The lemma trivially follows from (15), (16) and Assumption 3.1. In particular, note that for $x \in X$ and u a bounded measurable function on X , the function defined on $A(x)$ by

$$u'(a) := \int u(y) \widehat{Q}(dy | x, a)$$

can be written as

$$u'(a) = \frac{\tau}{\tau(x, a)} \int u(y) Q(dy | x, a) + u(x) \left(1 - \frac{\tau}{\tau(x, a)}\right). \blacksquare$$

The following two lemmas state that the associated MCM satisfies Assumptions 3.2 and 3.3.

LEMMA 5.3. *Suppose that Assumption 3.2 holds and for each $f \in \mathbf{F}$, let $\widehat{\phi}_f(x) := [\tau/\tau(x, f)]\phi_f(x)$. Then for all $B \in \mathcal{B}(X)$ and $x \in X$,*

- (a) $\widehat{Q}(B | x, f) \geq \widehat{\phi}_f(x)\nu(B)$.
- (b) *Under the additional condition (4),*

$$\int v(y) \widehat{Q}(dy | x, f) \leq \widehat{\phi}_f(x)\|v\|_v + \widehat{\alpha}v(x),$$

where $0 < \widehat{\alpha} < 1$ is the constant $\widehat{\alpha} := [1 - (\tau/M)(1 - \alpha)]$.

- (c) $\inf_{\mathbf{F}} \int \widehat{\phi}_f(y) \nu(dy) > 0$.

PROOF. Parts (a) and (c) follow directly from Assumption 3.2(a), 3.2(c) and from (16). On the other hand, by Assumption 3.2(b), defining $T(x, f) :=$

$\tau/\tau(x, f)$ we have

$$\begin{aligned} \int v(y) \widehat{Q}(dy | x, f) &= T(x, f) \int v(y) Q(dy | x, f) + [1 - T(x, f)]v(x) \\ &\leq T(x, f)\phi_f(x)\|\nu\|_v + T(x, f)\alpha v(x) + [1 - T(x, f)]v(x) \\ &\leq \widehat{\phi}_f(x)\|\nu\|_v + v(x)[1 - T(x, f)(1 - \alpha)] \\ &\leq \widehat{\phi}_f(x)\|\nu\|_v + \widehat{\alpha}v(x). \blacksquare \end{aligned}$$

LEMMA 5.4. *For each stationary policy f , the Markov chain defined by $\widehat{Q}(\cdot | \cdot, f)$ is ψ -irreducible, where ψ is the irreducibility measure in Assumption 3.3.*

PROOF. Choose an arbitrary stationary policy f , and let $T(x, f)$ be as in the proof of Lemma 5.3. Then, for all $x \in X$ and $B \in \mathcal{B}(X)$, we deduce from (16) that

$$(17) \quad \widehat{Q}(B | x, f) \geq T(x, f)Q(B | x, f) \geq \frac{\tau}{M}Q(B | x, f),$$

where M is the constant in (4). The latter inequality and Assumption 3.3 yield the desired conclusion. \blacksquare

As a consequence of Lemmas 5.3, 5.4, and 4.1 we obtain:

LEMMA 5.5. *For each stationary policy f , the Markov chain defined by $\widehat{Q}(\cdot | \cdot, f)$ is positive Harris recurrent. Hence, in particular, it admits a unique invariant probability measure \widehat{q}_f .*

We can also note in passing that $\widehat{Q}(\cdot | \cdot, f)$ is aperiodic since, by (16), $\widehat{Q}(\{x\} | x, f) \geq 1 - \tau/m > 0$ for each $x \in X$.

REMARK 5.6. From Lemmas 5.2 to 5.5 and 4.3, it follows that for each stationary policy $f \in \mathbf{F}$, there exists a pair $(\widehat{J}_f, \widehat{h}_f)$ with $\widehat{J}_f \in \mathbb{R}$ and $\widehat{h}_f \in L_v^\infty$ that is a solution to the Poisson equation for $\widehat{Q}(\cdot | \cdot, f)$, that is,

$$(18) \quad \widehat{h}_f(x) = \widehat{C}(x, f) - \widehat{J}_f + \int \widehat{h}_f(y) \widehat{Q}(dy | x, f)$$

for all $x \in X$. Therefore, integrating with respect to \widehat{q}_f (the invariant probability measure for \widehat{Q}) on both sides of (18), we get

$$\widehat{J}_f = \int \widehat{C}(y, f) \widehat{q}_f(dy) = \widehat{J}(f, x) := \lim_{n \rightarrow \infty} \frac{E_x^f \sum_{k=0}^{n-1} \widehat{C}(x_k, a_k)}{n}.$$

Now, defining $h_f(\cdot) := \tau \widehat{h}_f(\cdot)$, from (18) we obtain

$$h_f(x) = C(x, f) - \widehat{J}_f \tau(x, f) + \int h_f(y) Q(dy | x, f),$$

which [by Lemma 4.3 and (8)] implies

$$(19) \quad \widehat{J}_f = \widehat{J}(f, x) = J(f, x) = J_f.$$

6. Proof of Theorem 3.5. (a) We begin exactly as in [7], by showing the existence of a solution to the *optimality inequality* (20) below. However, in contrast to Theorem 2.6 in [7], in our present context the stationary policy f^* in (20) turns out to be optimal in \mathbf{F} only, not in all of Π [see paragraph (b) in Section 1]. More explicitly, there exist $j \in \mathbb{R}$, $\widehat{h} \in L_v^\infty$ and $f^* \in \mathbf{F}$ such that

(i) we have

$$(20) \quad \begin{aligned} \widehat{h}(x) &\geq \min_{a \in A(x)} \left\{ \widehat{C}(x, a) - j + \int \widehat{h}(y) \widehat{Q}(dy | x, a) \right\} \\ &= \widehat{C}(x, f^*) - j + \int \widehat{h}(y) Q(dy | x, f^*), \end{aligned}$$

(ii) $j = \widehat{J}(f^*, x)$ and f^* is AC-optimal on \mathbf{F} for the associated MCM, i.e.,

$$j = \widehat{J}(f^*, x) = \inf_{f \in \mathbf{F}} \widehat{J}(f, x) \quad \forall x \in X.$$

From the latter fact and (19) it follows that f^* is AC-optimal on \mathbf{F} for the original SMCM, i.e.,

$$j = J(f^*, x) = \inf_{f \in \mathbf{F}} J(f, x) \quad \forall x \in X.$$

To complete the proof of part (a), we shall now proceed as in the *policy iteration* (or *Howard's algorithm*) (cf. [9, 10, 16, 18, 24]).

Let $h'(x) := \tau \widehat{h}(x)$, where τ is the number in (16). If T is the operator defined on L_v^∞ by

$$Tu(x) := \min_{a \in A(x)} \left\{ C(x, a) - j\tau(x, a) + \int u(y) Q(dy | x, a) \right\},$$

then, from the inequality in (20), there exists $f_0 \in \mathbf{F}$ such that

$$h'(x) \geq Th'(x) = C(x, f_0) - j\tau(x, f_0) + \int h'(y) Q(dy | x, f_0).$$

This implies that $\widehat{J}(f^*, x) = j \geq J(f_0, x) = \widehat{J}(f_0, x) \geq \widehat{J}(f^*, x)$. Hence, $j = J(f_0, x)$ for all $x \in X$ and, therefore, also f_0 is AC-optimal in the class of stationary policies. On the other hand, taking $h_0 \in L_v^\infty$ satisfying [as in (10)]

$$(21) \quad h_0(x) = C(x, f_0) - j\tau(x, f_0) + \int h_0(y) Q(dy | x, f_0),$$

we deduce that the function $u(x) := h'(x) - h_0(x)$ is an f_0 -superharmonic function in L_v^∞ , that is, $u(x) \geq \int u(y) Q(dy | x, f_0)$ for all $x \in X$. This implies that $u(\cdot)$ is a constant q_{f_0} -a.e., and so

$$h'(x) = Th'(x) \quad q_{f_0}\text{-a.e.}$$

Now let f_1 be a stationary policy such that

$$(22) \quad Th_0(x) = C(x, f_1) - j\tau(x, f_1) + \int h_0(y) Q(dy | x, f_1),$$

and let (j_1, h_1) be a solution to the equation

$$(23) \quad h_1(x) = C(x, f_1) - j_1\tau(x, f_1) + \int h_1(y) Q(dy | x, f_1).$$

Then, from (21) and (22), we obtain

$$(24) \quad h_0(x) \geq C(x, f_1) - j\tau(x, f_1) + \int h_0(y) Q(dy | x, f_1),$$

and integrating both sides of (24) with respect to q_{f_1} we have

$$j \geq \frac{\int C(y, f_1) q_{f_1}(dy)}{\int \tau(y, f) q_{f_1}(dy)} = J(f_1, x) = j_1,$$

so that $j = j_1$. Moreover, subtracting (23) from (24),

$$h_0(x) - h_1(x) \geq \int [h_0(y) - h_1(y)] Q(dy | x, f_1),$$

which implies [by (7)]

$$h_0(x) - h_1(x) \geq \int [h_0(y) - h_1(y)] q_{f_1}(dy) \quad \forall x \in X,$$

and this in turn yields

$$h_0(x) - h_1(x) = \int [h_0(y) - h_1(y)] q_{f_1}(dy) \quad q_{f_1}\text{-a.e.}$$

Proceeding inductively we obtain sequences $\{f_n\} \subset \mathbf{F}$ and $\{h_n\} \subset L_v^\infty$ such that, for each n and $x \in X$, we have $j = J(f_n, x)$ and

$$h_n(x) = C(x, f_n) - j\tau(x, f_n) + \int h_n(y) Q(dy | x, f_n);$$

moreover,

$$(25) \quad Th_n(x) = C(x, f_{n+1}) - j\tau(x, f_{n+1}) + \int h_n(y) Q(dy | x, f_{n+1}),$$

and there exist measurable subsets N_{n+1} of X with $q_{f_{n+1}}(N_{n+1}) = 1$ such that

$$(26) \quad h_n(x) \begin{cases} = h_{n+1}(x) + k_{n+1} & \text{if } x \in N_{n+1}, \\ \geq h_{n+1}(x) + k_{n+1} & \text{if } x \in N_{n+1}^c, \end{cases}$$

where $k_{n+1} := \int [h_n(y) - h_{n+1}(y)] q_{f_{n+1}}(dy)$, $n = 0, 1, \dots$

Now, define $N := \bigcap_{n=1}^\infty N_n$ and observe that by Assumption 3.2(a), for each $B \in \mathcal{B}(X)$, $n = 1, 2, \dots$, and $f \in \mathbf{F}$,

$$Q^n(B | x, f) \geq \nu(B) \int \phi_f(y) Q^{n-1}(dy | x, f),$$

which yields [by (7)]

$$(27) \quad q_f(B) \geq \nu(B) \int \phi_f(y) q_f(dy).$$

Hence, since by Lemma 3.1(c) in [7],

$$\inf_{f \in \mathbf{F}} \int \phi_f(y) q_f(dy) > 0,$$

(27) implies that ν is absolutely continuous with respect to q_f for all $f \in \mathbf{F}$. Thus, $\nu(N_n) = 1$ for all $n \in \mathbb{N}$ and, therefore, $\nu(N) = 1$. Let $z \in N$ be an arbitrary, fixed state and define the functions

$$(28) \quad h_n^*(x) := h_n(x) - h_n(z), \quad x \in X, \quad n = 0, 1, \dots$$

Then, from (26),

$$(29) \quad h_n^*(\cdot) \geq h_{n+1}^*(\cdot) \quad \forall n = 0, 1, \dots$$

Define the function $h(x) := \lim_{n \rightarrow \infty} h_n^*(x)$, $x \in X$. Then [by (13)], $h \in L_v^\infty$ and we prove that $(j, h(\cdot))$ is a solution to the ACOE (5). Indeed, observe first that for all $x \in X$ and $n \in \mathbb{N}$,

$$Th(x) \leq Th_n^*(x) \leq C(x, f_n) - j\tau(x, f_n) + \int h_n^*(y) Q(dy | x, f_n) = h_n^*(x);$$

therefore, by (29),

$$(30) \quad Th(x) \leq h(x) \quad \forall x \in X.$$

On the other hand, combining (25) and (29) it follows that for all $x \in X$ and $n \in \mathbb{N}$,

$$\begin{aligned} Th_n^*(x) &= C(x, f_{n+1}) - j\tau(x, f_{n+1}) + \int h_n^*(y) Q(dy | x, f_{n+1}) \\ &\geq C(x, f_{n+1}) - j\tau(x, f_{n+1}) + \int h_{n+1}^*(y) Q(dy | x, f_{n+1}) \\ &= h_{n+1}^*(x) \geq h(x). \end{aligned}$$

As a consequence, for all $x \in X$ and $a \in A(x)$,

$$C(x, a) - j\tau(x, a) + \int h_n^*(y) Q(dy | x, a) \geq h(x).$$

Thus, letting $n \rightarrow \infty$, dominated convergence yields

$$C(x, a) - j\tau(x, a) + \int h(y) Q(dy | x, a) \geq h(x).$$

Therefore, $Th(x) \geq h(x)$ for all $x \in X$, and using (30) we conclude that $(j, h(\cdot))$ satisfies the ACOE, that is,

$$(31) \quad h(x) = \inf_{a \in A(x)} \left\{ C(x, a) - j\tau(x, a) + \int h(y) Q(dy | x, a) \right\} \quad \forall x \in X.$$

Finally, choosing $f \in \mathbf{F}$ such that $f(x)$ minimizes the right side of (31), we obtain (6), which completes the proof of part (a) of Theorem 3.5.

(b) Choose arbitrary $x \in X$ and $\pi \in \Pi$, and for each $k \in \mathbb{N}_0$, consider the corresponding history

$$i_k = (x_0, a_0, \delta_1, \dots, x_{k-1}, a_{k-1}, \delta_k, x_k),$$

and control $a_k \in A(x_k)$. Then

$$\begin{aligned} E_x^\pi [h(x_{k+1}) | (i_k, a_k)] &= \int h(y) Q(dy | x_k, a_k) \\ &= C(x_k, a_k) - j\tau(x_k, a_k) + \int h(y) Q(dy | x_k, a_k) \\ &\quad - C(x_k, a_k) + j\tau(x_k, a_k) \\ &\geq h(x_k) - C(x_k, a_k) + j\tau(x_k, a_k) \quad [\text{by (31)}]. \end{aligned}$$

Hence,

$$jE_x^\pi \sum_{k=0}^{n-1} \tau(x_k, a_k) \leq E_x^\pi h(x_n) - h(x) + E_x^\pi \sum_{k=0}^{n-1} C(x_k, a_k),$$

which yields

$$(32) \quad j + \frac{h(x)}{E_x^\pi \sum_{k=0}^{n-1} \tau(x_k, a_k)} \leq \frac{E_x^\pi \sum_{k=0}^{n-1} C(x_k, a_k) + E_x^\pi h(x_n)}{E_x^\pi \sum_{k=0}^{n-1} \tau(x_k, a_k)}.$$

Thus, taking $\liminf_{n \rightarrow \infty}$ and noting that $E_x^\pi \sum_{k=0}^{n-1} \tau(x_k, a_k) \geq nm$ [by (4)], from (12) we obtain

$$(33) \quad j \leq \liminf_{n \rightarrow \infty} \frac{E_x^\pi \sum_{k=0}^{n-1} C(x_k, a_k)}{E_x^\pi \sum_{k=0}^{n-1} \tau(x_k, a_k)} \leq J(\pi, x).$$

Therefore, since $j = J(f, x)$, and $\pi \in \Pi$ and $x \in X$ were arbitrary, part (b) follows from (33). ■

7. Example: a replacement model. In this section we present a semi-Markov replacement model that satisfies the hypotheses of Theorem 3.5.

A device is subjected to shocks which occur randomly in time. Each shock causes a random amount of damage and these damages accumulate additively. The time between two consecutive shocks is a random variable with distribution function $H(\cdot)$. Let $G(\cdot | x)$ be the conditional distribution of the magnitude of the shock associated with the state x . A failure can only happen at the time of a shock. If the accumulated damage is x and a shock of magnitude y occurs, then the device fails with a known probability $1 - r(x + y)$, where $r(\cdot)$ is a nonincreasing function called the *survival function*. When a device fails, it must be replaced by a new one and a failure cost is incurred; if the device is replaced before failure, a smaller (replacement) cost is incurred. The cost for operating the device is proportional to the accumulated damage.

The system is controlled in the following manner: at the time of a shock or after a replacement, one chooses the *planned replacement time* a , that is,

the time for the next replacement. Thus, a transition epoch is the time of a shock or the end of a planned replacement time.

This replacement model can be seen as a SMCM with the following components: $X = [0, \infty)$, and $A = A(x)$ for all x , where A is a compact subset of the interval $(0, M]$ for some finite number M . Furthermore, the transition law is

$$(34) \quad Q(B | x, a) := \delta_0(B) \left[1 - H(a) + H(a) \int_0^\infty (1 - r(x+y)) G(dy | x) \right] \\ + H(a) \int_{\{x+y \in B\}} r(x+y) G(dy | x),$$

with

$$F(t | x, a) := \begin{cases} H(t) & \text{if } t < a, \\ 1 & \text{if } t \geq a, \end{cases}$$

and we take a cost function of the form $C(x, a) := \beta x \tau(x, a) + KH(a) + k(1 - H(a))$, where β , k and K are positive constants with $k < K$.

Throughout the following we suppose:

ASSUMPTION 7.1. (a) $H(\cdot)$ is absolutely continuous with respect to the Lebesgue measure on \mathbb{R} , and $H(M) < 1$.

(b) There exists $q > 0$ such that for all $x \in X$,

$$(35) \quad \int_0^\infty e^{qy} r(x+y) G(dy | x) \leq 1.$$

[Note that a sufficient condition for (35) is $r(y) \leq e^{-qy}$ for all $y \in X$.]

The following propositions show that Assumption 7.1 implies Assumptions 3.1–3.3.

PROPOSITION 7.2. *Assumption 7.1(a) implies Assumption 3.1.*

PROOF. Assumption 3.1(a) follows from the definition of $A(x)$, whereas 3.1(c) is a consequence of the continuity of H on A . On the other hand, since

$$\tau(x, a) = \int_0^a z H(dz) + a(1 - H(a)),$$

Assumption 3.1(d) follows from the continuity of H and the fact that $0 < m := \inf A \leq \tau(x, a) \leq M$. To verify Assumption 3.1(b), consider the function $v(x) := e^{qx}$. Then, taking a suitable constant γ for which (3) holds, Assumption 3.1(b) follows from the continuity of H and the fact that for $(x, a) \in \mathbf{K}$,

$$\int v(y) Q(dy | x, a) = 1 - H(a) + H(a) \int_0^{\infty} [1 - r(x+y)] G(dy | x) \\ + H(a) \int_{\{y>0\}} e^{q(x+y)} r(x+y) G(dy | x). \blacksquare$$

PROPOSITION 7.3. *Assumption 7.1 implies Assumptions 3.2 and 3.3.*

PROOF. Let $\alpha := H(M)$, and $\nu(\cdot) := \delta_0(\cdot)$, the Dirac measure at $x = 0$. For $f \in \mathbf{F}$ define

$$\phi_f(x) := 1 - H(f(x)) + H(f(x)) \int_0^{\infty} [1 - r(x+y)] G(dy | x).$$

Then Assumption 3.2(a) follows immediately from (34). Moreover, since $\phi_f(x) \geq 1 - H(M)$ for all $x \in X$ and $f \in \mathbf{F}$, we obtain Assumption 3.2(c).

To verify Assumption 3.2(b), note that for any stationary policy $f \in \mathbf{F}$,

$$\int v(y) Q(dy | x, f) = \phi_f(x) \|\nu\|_v + H(f(x)) \int_{\{y>0\}} e^{q(x+y)} r(x+y) G(dy | x) \\ \leq \phi_f(x) \|\nu\|_v + H(M) e^{qx} \int_0^{\infty} e^{qy} r(x+y) G(dy | x),$$

and by Assumption 7.1(b), we obtain

$$\int v(y) Q(dy | x, f) \leq \phi_f(x) \|\nu\|_v + \alpha v(x).$$

Finally, using again the fact that $\phi_f(\cdot) \geq 1 - H(M) > 0$ and the inequality already proved in Assumption 3.2(a) we obtain Assumption 3.3 with $\psi := \nu$. \blacksquare

References

- [1] R. B. Ash, *Real Analysis and Probability*, Academic Press, New York, 1972.
- [2] R. N. Bhattacharya and M. Majumdar, *Controlled semi-Markov models under long-run average rewards*, J. Statist. Plann. Inference 22 (1989), 223–242.
- [3] 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), 222–235.
- [4] A. Federgruen, P. J. Schweitzer and H. C. Tijms, *Denumerable undiscounted semi-Markov decision processes with unbounded rewards*, Math. Oper. Res. 8 (1983), 298–313.
- [5] 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.
- [6] P. W. Glynn and S. P. Meyn, *A Lyapounov bound for solutions of Poisson's equations*, Ann. Probab. 24 (1996), 916–931.

- [7] 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.
- [8] U. G. Haussmann, *On the optimal long-run control of Markov renewal processes*, J. Math. Anal. Appl. 36 (1971), 123–140.
- [9] O. Hernández-Lerma and J. B. Lasserre, *Policy iteration for average cost Markov control processes on Borel spaces*, Acta Appl. Math. 47 (1997), 125–154.
- [10] —, —, *Further Topics on Discrete-Time Markov Control Processes*, Springer, New York, 1999 (in press).
- [11] M. Kurano, *Semi-Markov decision processes and their applications in the replacement models*, J. Oper. Res. Soc. Japan 28 (1985), 18–29.
- [12] S. A. Lippman, *Semi-Markov decision processes with unbounded rewards*, Management Sci. 19 (1973), 717–731.
- [13] —, *On dynamic programming with unbounded rewards*, *ibid.* 21 (1975), 1225–1233.
- [14] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer, London, 1993.
- [15] E. Nummelin, *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge Univ. Press, Cambridge, 1984.
- [16] M. L. Puterman, *Markov Decision Processes. Discrete Stochastic Dynamic Programming*, Wiley, New York, 1994.
- [17] U. Rieder, *Measurable selection theorems for optimization problems*, Manuscripta Math. 24 (1978), 115–131.
- [18] S. M. Ross, *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, 1970.
- [19] —, *Average cost semi-Markov decision processes*, J. Appl. Probab. 7 (1970), 649–656.
- [20] M. Schäl, *On the second optimality equation for semi-Markov decision models*, Math. Oper. Res. 17 (1992), 470–486.
- [21] P. J. Schweitzer, *Iterative solution of the functional equations of undiscounted Markov renewal programming*, J. Math. Anal. Appl. 34 (1971), 495–501.
- [22] L. I. Sennott, *Average cost semi-Markov decision processes and the control of queueing systems*, Probab. Engrg. Inform. Sci. 3 (1989), 247–272.
- [23] H. C. Tijms, *Stochastic Models: An Algorithmic Approach*, Wiley, Chichester, 1994.
- [24] O. Vega-Amaya, *Markov control processes in Borel spaces: undiscounted criteria*, doctoral thesis, UAM-Iztapalapa, México, 1998 (in Spanish).

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

Onésimo Hernández-Lerma
 Departamento de Matemáticas
 CINVESTAV-IPN, A. postal 14-740
 México D.F. 07000, México
 E-mail: ohernand@math.cinvestav.mx

*Received on 10.12.1998;
 revised version on 15.4.1999*