

O. HERNÁNDEZ-LERMA (México)  
O. VEGA-AMAYA (Sonora)

INFINITE-HORIZON MARKOV CONTROL PROCESSES  
WITH UNDISCOUNTED COST CRITERIA:  
FROM AVERAGE TO OVERTAKING OPTIMALITY

*Abstract.* We consider discrete-time Markov control processes on Borel spaces and infinite-horizon undiscounted cost criteria which are sensitive to the growth rate of finite-horizon costs. These criteria include, at one extreme, the grossly underselective *average cost* criterion and, at the other, the excessively overselective *strong overtaking optimality* introduced in the economics literature by Ramsey. Between these two we have the *overtaking* (or catching-up) *optimality* of Gale and von Weizsäcker, Flynn's *opportunity cost* and *F-strong average optimality*, Dutta's *D-strong average optimality*, Veinott's *bias-optimality*, and Yushkevich's *canonical policies*. In this paper we give conditions for the existence of optimal policies for each of these criteria, and show how they are interrelated, among other things.

**1. Introduction.** There are (deterministic and stochastic) control problems in engineering, economics and many others areas, in which it is not possible to specify *a priori* the optimization “horizon” or “termination time”, and one is led to consider infinite-horizon problems. These are roughly classified as *discounted* and *undiscounted*.

The former have a well-understood, unique meaning: at each decision time the running cost (or cost-per-stage) is multiplied by a discount factor, and then the control problem is to minimize the expected total *discounted cost* over the set of all admissible control policies.

The undiscounted case, on the other hand, is not that simple for, to begin with, it can mean different things. For instance, in the case in which the

---

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

*Key words and phrases*: (discrete-time) Markov control processes, undiscounted cost criteria, Lyapunov stability conditions, uniform ergodicity, Poisson's equation.

Research supported by CONACYT Grants 1332-E9206 and 3115P-E9608.

discount factor “vanishes” [see (11)–(12) below, with  $\alpha = 1$ ], the criterion to be minimized turns out to be the expected *total cost*, which is just the limit of finite-horizon costs as the horizon, say  $N$ , tends to infinity. This criterion, however, is not quite convenient in some applications [e.g., in queueing problems—see Ephremides and Verdú (1989), Stidham and Weber (1993)], one of the main “inconveniences” being that it does not take into account the rate at which the  $N$ -horizon costs vary as  $N \rightarrow \infty$ .

But then criteria which are sensitive to these cost rates—*which are precisely the criteria we are interested in*—can be defined in many different ways, from the grossly underselective *average cost* (AC) criterion to the extremely overselective *strong overtaking optimality* (strong OO) introduced by Ramsey (1928). Between these two extremes lie, for instance, the *overtaking optimality* (OO) of Gale (1967) and von Weizsäcker (1965); Flynn’s (1980) *opportunity cost* (OC-) and *F-strong average optimality* (F-strong AO); Dutta’s (1991) *D-strong average optimality* (D-strong AO); Veinott’s (1966) *bias-optimality*; and Yushkevich’s (1973) *canonical policies*.

In this paper we consider discrete-time *Markov control processes* (MCPs) with *Borel state and action spaces*, and the main objective is to give conditions for the existence of optimal policies for each of the above-mentioned criteria and to study how these criteria are interrelated. For instance, under appropriate hypotheses, one of our main results (Theorem 3.5) states the existence of canonical policies and that, moreover, “canonical”, “F-strong AO” and “AC-optimal” are all equivalent notions. Similarly, we show the existence of “D-strong AO” policies in the class of “stationary policies” (see Theorem 4.9), and that the latter are equivalent to “bias-optimal”, “OO” and “OC-optimal” policies (Theorem 4.12).

The remainder of the paper is organized into six sections. In Section 2 we introduce the MCP we will be dealing with, together with our hypotheses (Assumptions 2.4, 2.5 and 2.8) and other relevant information. Sections 3 and 4 present our main results on AC-like and growth-sensitive criteria, respectively; Section 4 also includes an example showing that without appropriate assumptions, some of the main results may not be true. The proofs of all our results are given in Section 5. In Section 6 we present an example that illustrates how to verify our assumptions. The paper concludes in Section 7 with some general remarks and a brief discussion of some related open problems.

REMARK 1.1. The reader should be warned that not everyone uses the same definitions for the several optimality criteria presented here. For instance, what we call “strong overtaking optimality” is referred to as “overtaking optimality (OO)” by Fernández-Gaucherand *et al.* (1994), and our

“OO” is sometimes called “catching-up” in the economics literature, for instance, in Dutta (1991) and Gale (1967).

REMARK 1.2. *Terminology and notation.* A Borel subset of a complete and separable metric space is called a *Borel space*. A topological space, say  $X$ , is always endowed with the Borel  $\sigma$ -algebra  $\mathcal{B}_X$ . If  $X$  and  $Y$  are Borel spaces, a *stochastic kernel on  $X$  given  $Y$*  is a function  $P(B|y)$  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 fixed  $B \in \mathcal{B}_X$ . We denote by  $\mathbb{N}$  (respectively  $\mathbb{N}_0$ ) the set of positive (resp. nonnegative) integers. Other terminology and notation we use is quite standard in the MCP literature: see Araposthatis *et al.* (1993), Bertsekas and Shreve (1978), Dynkin and Yushkevich (1979), Hernández-Lerma (1989), Hernández-Lerma and Lasserre (1996), etc.

**2. Preliminaries.** We consider a (discrete-time) *Markov control model*  $(\mathbf{X}, \mathbf{A}, \{\mathbf{A}(x) : x \in \mathbf{X}\}, Q, C)$  satisfying the following conditions. The *state space*  $\mathbf{X}$  and the *control (or action) set*  $\mathbf{A}$  are both Borel spaces and, in addition,  $\mathbf{X}$  is locally compact. For every state  $x \in \mathbf{X}$ ,  $\mathbf{A}(x)$  is a non-empty Borel subset of  $\mathbf{A}$ ; the elements of  $\mathbf{A}(x)$  are the *feasible actions* if the system is in the state  $x$ . The set  $\mathbf{K} := \{(x, a) : x \in \mathbf{X}, a \in \mathbf{A}(x)\}$  of *feasible state-action pairs* is assumed to be a Borel subset of  $\mathbf{X} \times \mathbf{A}$  and to contain the graph of a measurable map from  $\mathbf{X}$  to  $\mathbf{A}$ . Finally, the *transition law*  $Q$  is a stochastic kernel on  $\mathbf{X}$  given  $\mathbf{K}$ , and the *one-stage cost*  $C$  is a real-valued measurable function on  $\mathbf{K}$ .

For every  $n \in \mathbb{N}_0$ , let  $\mathbf{H}_n$  be the family of admissible histories up to time  $n$ ; that is,  $\mathbf{H}_0 := \mathbf{X}$ , and  $\mathbf{H}_n := \mathbf{K} \times \mathbf{H}_{n-1}$  if  $n \geq 1$ .

DEFINITION 2.1. A (*randomized*) *policy* is a sequence  $\pi = \{\pi_n\}$  of stochastic kernels  $\pi_n$  on  $\mathbf{A}$  given  $\mathbf{H}_n$  such that  $\pi_n(\mathbf{A}(x_n)|h_n) = 1$  for every  $n$ -history  $h_n = (x_0, a_0, \dots, x_{n-1}, a_{n-1}, x_n)$  in  $\mathbf{H}_n$ . We denote by  $\mathbf{\Pi}$  the set of all policies. A *deterministic policy* is a sequence  $\{f_n\}$  of measurable maps  $f : \mathbf{H}_n \rightarrow \mathbf{A}$  such that  $f_n(h_n) \in \mathbf{A}(x_n)$  for all  $h_n \in \mathbf{H}_n$ . [A deterministic policy  $\{f_n\}$  can be identified with the randomized policy  $\pi = \{\pi_n\}$  such that  $\pi_n(\cdot|h_n)$  is the probability measure concentrated at  $f_n(h_n)$  for all  $n \in \mathbb{N}_0$ ]. Let  $\mathbf{F}$  be the set of all *decision functions*, i.e., measurable functions  $f : \mathbf{X} \rightarrow \mathbf{A}$  such that  $f(x) \in \mathbf{A}(x)$  for all  $x \in \mathbf{X}$ . Then a sequence  $\{f_n\}$  of functions in  $\mathbf{F}$  is called a (*deterministic*) *Markov policy*. Finally, a (*deterministic*) *stationary policy*  $\{f_n\}$  is a Markov policy such that  $f_n \equiv f$  is independent of  $n$ . We will identify  $\mathbf{F}$  with the set of all stationary policies.

REMARK 2.2. (a) Every initial state  $x$  and policy  $\pi$  induce a probability measure  $P_x^\pi$  on the measurable space  $(\Omega, \mathcal{G})$ , where  $\Omega := (\mathbf{X} \times \mathbf{A})^\infty$  and  $\mathcal{G}$

is the corresponding product  $\sigma$ -algebra; the expectation operator is denoted by  $E_x^\pi$  [see Bertsekas and Shreve (1978), Dynkin and Yushkevich (1979), ...].

(b) For a stationary policy (or decision function)  $f \in \mathbf{F}$  we write

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

To state the assumptions used in this paper, let us first introduce the following definition [see Meyn and Tweedie (1993), Glynn and Meyn (1996)].

DEFINITION 2.3. Let  $\mathbf{V}(\cdot) \geq 1$  be a given measurable function on  $\mathbf{X}$ , and let  $\mathbf{L}_\mathbf{V}^\infty$  be the normed linear space of all measurable functions  $u : \mathbf{X} \rightarrow \mathbb{R}$  with

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

For every  $u$  in  $\mathbf{L}_\mathbf{V}^\infty$  define a new function  $\mathbf{T}u$  on  $\mathbf{X}$  by

$$(1) \quad \mathbf{T}u(x) := \inf_{a \in \mathbf{A}(x)} \left[ C(x, a) + \int_{\mathbf{X}} u(y) Q(dy | x, a) \right], \quad x \in \mathbf{X}.$$

The following Assumptions 2.4 and 2.5 guarantee that, *inter alia*, for every  $x \in \mathbf{X}$ , the minimum is actually attained on the right-hand side of (1)—in which case we write “min” instead of “inf”—and that, moreover,  $\mathbf{T}$  maps  $\mathbf{L}_\mathbf{V}^\infty$  into itself.

ASSUMPTION 2.4. For every  $x \in \mathbf{X}$ :

- (a)  $\mathbf{A}(x)$  is a compact subset of  $\mathbf{A}$ ;
- (b)  $C(x, \cdot)$  is l.s.c. (lower semicontinuous) on  $\mathbf{A}(x)$ ;
- (c)  $\int_{\mathbf{X}} u(y) Q(dy | x, a)$  is continuous in  $a \in \mathbf{A}(x)$  for every bounded and measurable function  $u$  on  $\mathbf{X}$ ;
- (d)  $\int_{\mathbf{X}} \mathbf{V}(y) Q(dy | x, a)$  is continuous in  $a \in \mathbf{A}(x)$ , where  $\mathbf{V}$  is the function in Definition 2.3;
- (e)  $\sup_{a \in \mathbf{A}(x)} |C(x, a)| \leq \mathbf{V}(x)$ .

We shall next require the function  $\mathbf{V}$  in Assumption 2.4 to satisfy a Lyapunov stability condition as in Glynn and Meyn (1996), or Meyn and Tweedie (1993), p. 367. In the statement of Assumption 2.5 we use the terminology from the latter references.

ASSUMPTION 2.5. (a) The function  $\mathbf{V}$  in Assumption 2.4 satisfies the following: for each  $f \in \mathbf{F}$  there is a petite set  $C_f$  and constants  $B_f < 1$  and  $b_f < \infty$  such that, for all  $x \in \mathbf{X}$  [using the notation of Remark 2.2(b)],

$$(2) \quad \int_{\mathbf{X}} \mathbf{V}(y) Q(dy | x, f) \leq B_f \mathbf{V}(x) + b_f 1_{C_f}(x),$$

where  $1_{C_f}$  denotes the indicator function of  $C_f$ .

(b) For every stationary policy  $f$  the corresponding Markov (state) process  $\{x_t\}$  is  $\psi$ -irreducible, for some  $\sigma$ -finite measure  $\psi$  (independent of  $f$ ) on  $\mathbf{X}$ , and aperiodic.

PROPOSITION 2.6. *Assumptions 2.4 and 2.5(a) imply the following:*

(a) *For every  $u$  in  $\mathbf{L}_{\mathbf{V}}^{\infty}$  there exists a decision function  $f$  (depending of  $u$ ) such that*

$$(3) \quad \mathbf{T}u(x) = C(x, f) + \int_{\mathbf{X}} u(y) Q(dy | x, f) \quad \text{for all } x \in \mathbf{X}.$$

(b)  *$\mathbf{T}$  maps  $\mathbf{L}_{\mathbf{V}}^{\infty}$  into itself, i.e., if  $u$  is in  $\mathbf{L}_{\mathbf{V}}^{\infty}$ , then so is  $\mathbf{T}u$ .*

(c) *There exist constants  $B < 1$  and  $b < \infty$ , and a subset  $C$  of  $\mathbf{X}$  (which is petite with respect to  $Q(\cdot | x, g)$  for some decision function  $g$ ) such that*

$$(4) \quad \int_{\mathbf{X}} \mathbf{V}(y) Q(dy | x, a) \leq B\mathbf{V}(x) + b1_C(x) \quad \text{for all } (x, a) \in \mathbf{K}.$$

As shown by Glynn and Meyn (1996) [see also Meyn and Tweedie (1993)], Assumption 2.5 has the following important consequences (which are used in the proof of our main results):

THEOREM 2.7. *Assumption 2.5 implies that for every stationary policy  $f$ :*

(a)  *$\{x_t\}$  is a positive Harris-recurrent chain with a unique invariant probability measure  $Q_f$ .*

(b)  *$\int_{\mathbf{X}} \mathbf{V} dQ_f < \infty$ .*

(c)  *$\{x_t\}$  is  $\mathbf{V}$ -uniformly ergodic, i.e., there exist positive constants  $\gamma_f < 1$  and  $M_f < \infty$  such that*

$$(5) \quad \left| \int_{\mathbf{X}} u(y) Q^n(dy | x, f) - \int_{\mathbf{X}} u dQ_f \right| \leq \|u\|_{\mathbf{V}} M_f \gamma_f^n \mathbf{V}(x)$$

for all  $x \in \mathbf{X}$ ,  $n \in \mathbb{N}_0$ , and  $u$  in  $\mathbf{L}_{\mathbf{V}}^{\infty}$ .

(d) *There exists a function  $h_f$  in  $\mathbf{L}_{\mathbf{V}}^{\infty}$  such that the pair  $(J(f), h_f)$ , with  $J(f) := \int_{\mathbf{X}} C(y, f) Q_f(dy)$ , satisfies the Poisson equation*

$$(6) \quad J(f) + h_f(x) = C(x, f) + \int_{\mathbf{X}} h_f(y) Q(dy | x, f) \quad \text{for all } x \in \mathbf{X};$$

moreover,  $h_f$  can be defined as

$$(7) \quad h_f(x) := \sum_{t=0}^{\infty} E_x^f [C(x_t, f) - J(f)], \quad x \in \mathbf{X}.$$

Our final assumption concerns the constants  $\gamma_f$  and  $M_f$  in (5).

ASSUMPTION 2.8.  $\mathbf{M} := \sup_{\mathbf{F}} M_f$  and  $\gamma := \sup_{\mathbf{F}} \gamma_f$  satisfy

$$(8) \quad \mathbf{M} < \infty \quad \text{and} \quad \gamma < 1.$$

For examples of control models in which Assumptions 2.4, 2.5 and 2.8 hold, see, for instance, Gordienko and Hernández-Lerma (1995b) and Vega-Amaya (1996).

The key ergodicity-related facts we use in our proofs below are (5), (6) and (8). Alternative approaches to obtain these—for MCPs with *bounded* costs—appear, for instance, in Hernández-Lerma (1989), Hernández-Lerma, Montes-de-Oca and Cavazos-Cadena (1991), Nowak (1992), and the papers cited there. See also Remark 2.10 below.

REMARK 2.9. (a) Let  $\psi$  and  $Q_f$  ( $f \in \mathbf{F}$ ) be as in Assumption 2.5(b) and Theorem 2.7(a) respectively. We shall occasionally use the fact that, for every  $f \in \mathbf{F}$ ,  $\psi$  is absolutely continuous with respect to  $Q_f$ , i.e.,  $Q_f(B) = 0$  implies  $\psi(B) = 0$  [Meyn and Tweedie (1993), Proposition 10.1.2; Orey (1971), Theorem 7.2]. Moreover, we shall consider two functions in  $\mathbf{L}_V^\infty$  to be  $Q_f$ -*equivalent* if they are equal  $Q_f$ -almost everywhere ( $Q_f$ -a.e.); moreover, for each fixed decision function  $f$ , we do not distinguish between  $Q_f$ -equivalent functions.

(b) For any stationary policy  $f$ , if  $h_1$  and  $h_2$  are two functions in  $\mathbf{L}_V^\infty$  that satisfy the Poisson equation (6), then  $h_1$  and  $h_2$  coincide except perhaps for an additive constant; in fact,

$$(9) \quad h_1(x) - h_2(x) = \int_{\mathbf{X}} [h_1(y) - h_2(y)] Q_f(dy) \quad \text{for all } x \in \mathbf{X}.$$

Indeed, if for all  $x \in \mathbf{X}$ ,

$$J(f) + h_i(x) = C(x, f) + \int_{\mathbf{X}} h_i(y) Q(dy | x, f), \quad i = 1, 2,$$

then  $u := h_1 - h_2$  satisfies

$$u(x) = \int_{\mathbf{X}} u(y) Q(dy | x, f).$$

Therefore  $u(x) = E_x^f u(x_n)$  for all  $x \in \mathbf{X}$  and  $n \in \mathbb{N}_0$ , which [letting  $n \rightarrow \infty$  and using (5)] yields (9). The last fact we used can be explicitly stated as:

(c) By (5),  $\lim_{n \rightarrow \infty} E_x^f u(x_n) = \int_{\mathbf{X}} u dQ_f$  for all  $f \in \mathbf{F}$  and  $u \in \mathbf{L}_V^\infty$ .

(d) The following facts will be used repeatedly in the proofs of our results: from Proposition 2.6(c), we have, for all  $u \in \mathbf{L}_V^\infty$  and  $x \in \mathbf{X}$ ,

$$\limsup_{n \rightarrow \infty} \sup_{\pi \in \Pi} E_x^\pi |u(x_n)| \leq \|u\|_{\mathbf{V}} b(1 - B)^{-1},$$

which in turn implies

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sup_{\pi \in \Pi} E_x^\pi |u(x_n)| = 0.$$

This follows directly from the inequality  $E_x^\pi |u(x_n)| \leq \|u\|_{\mathbf{V}} E_x^\pi \mathbf{V}(x_n)$  and

an inductive argument using (4), which yields

$$E_x^\pi |u(x_n)| \leq \|u\|_{\mathbf{V}} \left[ B^n \mathbf{V}(x) + b \sum_{k=0}^{n-1} B^k \right].$$

REMARK 2.10. An important particular case of Assumption 2.5, especially when one deals with *bounded* cost functions, is the case in which  $\mathbf{V}(\cdot)$  is a bounded function. In such a situation,  $\mathbf{L}_{\mathbf{V}}^\infty$  becomes the space of *bounded* measurable functions on  $\mathbf{X}$ , and the  $\mathbf{V}$ -uniform ergodicity (5) reduces to the usual *uniform ergodicity* condition

$$\|Q^n(\cdot | x, f) - Q_f(\cdot)\|_{\text{TV}} \leq M_f \gamma_f^n,$$

where  $\|\cdot\|_{\text{TV}}$  denotes the *total variation norm* for finite signed measures. Moreover, for every stationary policy  $f \in \mathbf{F}$ , the following conditions are equivalent [Meyn and Tweedie (1993), Theorem 16.0.2]:

- (a)  $\{x_t\}$  is uniformly ergodic;
- (b)  $\{x_t\}$  is aperiodic and there is a *bounded* solution  $\mathbf{V} \geq 1$  to the inequality (2);
- (c)  $\{x_t\}$  is aperiodic and *Doebelin's condition* holds; that is, there exists a probability measure  $\lambda_f$  on  $\mathcal{B}_{\mathbf{X}}$ , positive numbers  $\varepsilon < 1$  and  $\delta$ , and an integer  $m \geq 1$  such that

$$\inf_{x \in \mathbf{X}} Q^m(B|x, f) > \delta \quad \text{whenever} \quad \lambda_f(B) > \varepsilon.$$

**3. Average-cost optimality criteria.** For every  $n \in \mathbb{N}$ , let

$$(10) \quad J_n(\pi, x) := E_x^\pi \sum_{t=0}^{n-1} C(x_t, a_t)$$

be the *expected total cost* when using the policy  $\pi$ , given the initial state  $x_0 = x$ . As already mentioned in the introduction, the basic problem is how to define suitable optimality criteria in the infinite-horizon case,  $n = \infty$ . A standard way of doing this is to introduce a *discount factor*  $\alpha$ , with  $\alpha \in (0, 1)$ , and define the infinite-horizon  $\alpha$ -*discounted cost*

$$(11) \quad V_\alpha(\pi, x) := \lim_{n \rightarrow \infty} E_x^\pi \sum_{t=0}^{n-1} \alpha^t C(x_t, a_t).$$

Although this criterion will be useful in later sections, the main concern in this paper is the case of *undiscounted* ( $\alpha = 1$ ) cost criteria. Thus we might consider the infinite-horizon *expected total cost*

$$(12) \quad V(\pi, x) := \lim_{n \rightarrow \infty} J_n(\pi, x).$$

This, however, is not quite what we are interested in because for many applications  $J_n(\pi, x)$  diverges for some or all policies, or the limit in (12)

may not exist [as, say, in the examples in Puterman (1994), Section 5.1]. It turns out to be more useful to look for optimality criteria which are sensitive to the rate at which  $J_n(\pi, x)$  varies as  $n \rightarrow \infty$ .

One such criterion is the long-run *expected average cost* (AC) [perhaps introduced by Bellman (1957)]

$$(13) \quad J(\pi, x) := \limsup_{n \rightarrow \infty} \frac{1}{n} J_n(\pi, x),$$

with corresponding *AC-value function*

$$(14) \quad J^*(x) := \inf_{\Pi} J(\pi, x), \quad x \in \mathbf{X}.$$

DEFINITION 3.1. A policy  $\pi^*$  is said to be *average cost optimal* (AC-optimal) if

$$(15) \quad J^*(x) = J(\pi^*, x) \quad \text{for all } x \in \mathbf{X}.$$

AC-optimality is a widely used criterion in many applications [for instance, in the analysis of communication networks and queueing systems: Ephremides and Verdú (1989), Stidham and Weber (1993), ...], but it has the disadvantage of being grossly *underselective*, i.e., the AC-criterion may not distinguish between policies which have quite different finite-horizon behavior. At the other extreme we have the excessively *overselective* criterion—introduced by Ramsey (1928)—of *strong overtaking optimality*, which will be discussed in Section 4.

We next define related optimality criteria using the AC-value function in (14) and the *n-stage value function*

$$(16) \quad J_n^*(x) := \inf_{\Pi} J_n(\pi, x), \quad x \in \mathbf{X},$$

with  $J_n(\pi, x)$  as in (10).

DEFINITION 3.2. A policy  $\pi^*$  is said to be *F-strong average optimal* (abbreviated F-strong AO or F-SAO)—“F” for Flynn (1980)—if

$$(17) \quad \lim_{n \rightarrow \infty} \frac{1}{n} [J_n(\pi^*, x) - J_n^*(x)] = 0 \quad \text{for all } x \in \mathbf{X}.$$

REMARK 3.3. (a) Note that for any policy  $\pi$  and initial state  $x$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} J_n^*(x) \leq J(\pi, x),$$

which implies [by (14)]

$$\limsup_{n \rightarrow \infty} \frac{1}{n} J_n^*(x) \leq J^*(x).$$

(b) If  $\pi^*$  is F-SAO, then it is AC-optimal. In fact, from (a) and (17),

$$\limsup_{n \rightarrow \infty} \frac{1}{n} J_n^*(x) = J(\pi^*, x) = J^*(x) \quad \text{for all } x \in \mathbf{X}.$$



We shall now introduce the notion of *canonical triplets* [cf. Yushkevich (1973)], which requires the following terminology. Let  $h : \mathbf{X} \rightarrow \mathbb{R}$  be a given measurable function, and define the *n-stage expected total cost*, with *terminal cost function*  $h$ , as

$$(18) \quad J_n(\pi, x, h) := E_x^\pi \left[ \sum_{t=0}^{n-1} C(x_t, a_t) + h(x_n) \right], \quad n \in \mathbb{N}, \quad x \in \mathbf{X},$$

with corresponding value function

$$(19) \quad J_n^*(x, h) := \inf_{\Pi} J_n(\pi, x, h), \quad x \in \mathbf{X}.$$

DEFINITION 3.4. Let  $\varrho$  and  $h$  be two given real-valued functions on  $\mathbf{X}$ , and let  $f$  be a stationary policy. Then  $(\varrho, h, f)$  is said to be a *canonical triplet* if

$$(20) \quad J_n(f, x, h) = J_n^*(x, h) = n\varrho(x) + h(x) \quad \text{for all } x \in \mathbf{X}, \quad n \in \mathbb{N}_0.$$

A stationary policy is said to be *canonical* if it enters into some canonical triplet.

One of the main objectives of this paper is to show the existence of a canonical triplet  $(\varrho, h, f)$  with  $\varrho$  a constant function, say  $\varrho(\cdot) \equiv \varrho^*$ . This is more precisely stated in part (a) of the following theorem, where it is also shown that, under the assumptions given in Section 2, for a policy to be canonical, F-SAO and AC-optimal are all equivalent concepts.

THEOREM 3.5. *Suppose that Assumptions 2.4, 2.5 and 2.8 hold. Then:*

(a) *There exists a constant  $\varrho^*$ , a function  $h$  in  $\mathbf{L}_V^\infty$  and a stationary policy  $f$  such that  $(\varrho^*, h, f)$  is a canonical triplet, i.e.,*

$$(21) \quad J_n(f, x, h) = J_n^*(x, h) = n\varrho^* + h(x) \quad \text{for all } x \in \mathbf{X}, \quad n \in \mathbb{N}_0.$$

*Equivalently,  $(\varrho^*, h, f)$  satisfies the Average Cost Optimality Equation (ACOE)*

$$(22) \quad \begin{aligned} \varrho^* + h(x) &= \min_{a \in A(x)} \left[ C(x, a) + \int_{\mathbf{X}} h(y) Q(dy | x, a) \right] \\ &= C(x, f) + \int_{\mathbf{X}} h(y) Q(dy | x, f). \end{aligned}$$

*Moreover,  $f$  is AC-optimal and the AC-value function is  $\varrho^*$ , i.e.,*

$$(23) \quad J(f, x) = J^*(x) = \varrho^* \quad \text{for all } x \in \mathbf{X}.$$

(b) *The following statements are equivalent:*

- (i)  *$f$  is AC-optimal;*
- (ii)  *$f$  is a canonical policy;*
- (iii)  *$f$  is F-SAO.*

REMARK 3.6. (a) The equivalence of (21) and (22) is well known: for a proof in the case of a *bounded* one-stage cost function see, for instance, Yushkevich (1973), Dynkin and Yushkevich (1979), or Arapostathis *et al.* (1993); for general  $C$ , a proof may be found in Montes-de-Oca and Hernández-Lerma (1996) or Hernández-Lerma and Lasserre (1996), Section 5.2.

(b) From the results in the last two references [or in Gordienko and Hernández-Lerma (1995a)] it can be seen that the statements in Theorem 3.5(b) are also equivalent to:  $f \in \mathbf{F}$  is *asymptotically optimal*, by which we mean the following [see Dynkin and Yushkevich (1979)]:

$$J(f, x) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} J_n(\pi, x) \quad \text{for all } \pi \in \mathbf{\Pi}, x \in \mathbf{X}.$$

(c) Note that if  $h_1$  is a function that solves the ACOE (22), then so does  $h_1 + c$  for any real constant  $c$ . Thus we can take the function  $h$  in Theorem 3.5(a) satisfying  $h(z) = 0$  where  $z$  is an arbitrary but fixed state. Conversely:

(d) The same argument used in the proof of parts (a)–(b) of Theorem 3.5 [see (55) and (57)] can be used to show that if  $h_1$  and  $h_2$  are two functions in  $\mathbf{L}_V^\infty$  that satisfy the ACOE, i.e., for all  $x \in \mathbf{X}$ ,

$$\varrho^* + h_i(x) = \min_{a \in A(x)} \left[ C(x, a) + \int_{\mathbf{X}} h_i(y) Q(dy | x, a) \right], \quad i = 1, 2,$$

then  $h_1$  and  $h_2$  are the same function except perhaps for an additive constant.

**4. Growth-sensitive criteria.** In this section we discuss several criteria which are “sensitive” to the growth rate of the finite-horizon costs. We begin with the excessively overselective criterion of *strong overtaking optimality*.

DEFINITION 4.1. (a) A policy  $\pi^*$  is said to *overtake* a policy  $\pi$  if for every initial state  $x$  there exists an  $N = N(\pi^*, \pi, x)$  such that

$$(24) \quad J_n(\pi^*, x) \leq J_n(\pi, x) \quad \text{for all } n \geq N.$$

(b)  $\pi^*$  is called *strong overtaking optimal* (strong OO) if it overtakes any other policy  $\pi$ .

Note that if both sequences in (24) converge, then overtaking is equivalent to the comparison of  $\pi^*$  and  $\pi$  according to the total cost criterion (12). Moreover, strong overtaking optimality is obviously stronger than AC-optimality. There are, on the other hand, many well-known, elementary examples showing that strong OO policies need not exist [Brown (1965), Puterman (1994)]. In fact, in many applications this criterion is bound to be very restrictive to be useful; thus we consider the following weaker criteria:

DEFINITION 4.2 [Gale (1967), von Weizsäcker (1965)]. A policy  $\pi^*$  is said to be *overtaking optimal* (OO) if for every policy  $\pi$ , initial state  $x$ , and  $\varepsilon > 0$  there exists an  $N = N(\pi^*, \pi, x, \varepsilon)$  such that

$$(25) \quad J_n(\pi^*, x) \leq J_n(\pi, x) + \varepsilon \quad \text{for all } n \geq N.$$

DEFINITION 4.3. (a) [Flynn (1980)] Let

$$(26) \quad \text{OC}(\pi, x) := \limsup_{n \rightarrow \infty} [J_n(\pi, x) - J_n^*(x)]$$

be the *opportunity cost* (OC) of the policy  $\pi$  given the initial state  $x$ , and define the *OC-value function* as

$$(27) \quad \text{OC}^*(x) := \inf_{\Pi} \text{OC}(\pi, x), \quad x \in \mathbf{X}.$$

A policy  $\pi^*$  is said to be *opportunity cost optimal* (OC-optimal) if

$$(28) \quad \text{OC}(\pi^*, x) = \text{OC}^*(x) \quad \text{for all } x \in \mathbf{X}.$$

(b) [Dutta (1991)] Given a policy  $\pi$  and the initial state  $x$ , let [with  $J^*$  as in (14)]

$$(29) \quad D^u(\pi, x) := \limsup_{n \rightarrow \infty} [J_n(\pi, x) - nJ^*(x)],$$

$$(30) \quad D^l(\pi, x) := \liminf_{n \rightarrow \infty} [J_n(\pi, x) - nJ^*(x)]$$

be the *upper D-strong average cost* (“D” for Dutta) and the *lower D-strong average cost* of  $\pi$  at  $x$ , respectively. Then a policy  $\pi^*$  is said to be *D-strong average optimal* (D-strong AO) if

$$(31) \quad D^u(\pi^*, x) = \inf_{\Pi} D^u(\pi, x) \quad \text{for all } x \in \mathbf{X}.$$

Moreover, if for all  $x \in \mathbf{X}$ ,

$$(32) \quad \inf_{\Pi} D^u(\pi, x) = \inf_{\Pi} D^l(\pi, x) =: D^*(x),$$

then  $D^*$  is called the *D-strong average value function*.

REMARK 4.4. Define the upper and lower limit functions

$$(33) \quad L^u(x) := \limsup_{n \rightarrow \infty} [J_n^*(x) - nJ^*(x)],$$

$$(34) \quad L^l(x) := \liminf_{n \rightarrow \infty} [J_n^*(x) - nJ^*(x)].$$

A straightforward calculation shows that for any policy  $\pi$  and initial state  $x$  for which  $|D^u(\pi, x)| < \infty$ , the opportunity cost and the upper D-strong average cost—see (26) and (29)—are related as follows:

$$(35) \quad D^u(\pi, x) - L^u(x) \leq \text{OC}(\pi, x) \leq D^u(\pi, x) - L^l(x).$$

The following proposition summarizes some basic relations between the optimality criteria introduced in this section. Its proof is omitted since it follows from direct arguments.

PROPOSITION 4.5. (a) *If  $\pi^*$  is OO, then  $\pi^*$  is both D-strong AO and OC-optimal. Conversely:*

(b) *If  $\pi^*$  is D-strong AO and the equality in (32) holds, then  $\pi^*$  is OO [hence  $\pi^*$  is OC-optimal by (a)]. Similarly, if  $\pi^*$  is OC-optimal and*

$$(36) \quad \inf_{\Pi} \text{OC}(\pi, x) = \inf_{\Pi} \text{OC}^l(\pi, x) \quad \text{for all } x \in \mathbf{X},$$

where

$$(37) \quad \text{OC}^l(\pi, x) := \liminf_{n \rightarrow \infty} [J_n(\pi, x) - J_n^*(x)],$$

then  $\pi^*$  is OO [hence  $\pi^*$  is D-strong AO];

(c) *If  $\text{OC}(\pi, \cdot) < \infty$ , then  $\pi$  is F-SAO. Similarly, if  $D^u(\pi, \cdot) < \infty$  and  $L^l(\cdot) > -\infty$ , then  $\pi$  is F-SAO.*

(d) *If  $\pi$  is not AC-optimal at  $x$  [i.e.,  $J(\pi, x) > J^*(x)$ ] then*

$$D^u(\pi, x) = \infty \quad \text{and} \quad \text{OC}(\pi, x) = \infty.$$

REMARK 4.6. (a) Condition (36) holds if, for instance,  $\text{OC}^*(x) = 0$  for all states  $x$ . This follows from the fact that  $\text{OC}(\pi, x) \geq \text{OC}^l(\pi, x) \geq 0$ .

(b) From Proposition 4.5(a)–(b), if (32) and (36) both hold, then the concepts of overtaking optimality (OO), D-strong average optimality (D-strong AO) and opportunity cost optimality (OC-optimality) are all equivalent. In Theorem 4.12 we show that this equivalence also holds if we restrict ourselves to work with stationary policies.

(c) It follows from Proposition 4.5(a), (d) that when dealing with OO, D-strong AO and AC-optimality one can essentially restrict the analysis to average cost (AC-) optimal policies. In the terminology of Gale (1967) and Dutta (1991), a policy  $\pi$  for which  $D^u(\pi, x) < \infty$  is said to be a “good” policy.

For every stationary policy  $f$ , let  $(J(f), h_f)$  be as in Theorem 2.7(d), i.e., a solution to the Poisson equation (6), and define

$$(38) \quad \bar{h}_f(x) := h_f(x) - \int_{\mathbf{X}} h_f(y) Q_f(dy), \quad x \in \mathbf{X},$$

$$(39) \quad \bar{h}(x) = \inf\{\bar{h}_f(x) : f \in \mathbf{F} \text{ and } J(f) = \varrho^*\}, \quad x \in \mathbf{X},$$

where  $\varrho^*$  is the (constant) AC-value function in (23). Note that, by Theorem 3.5(a), we may also write  $\varrho^*$  as

$$(40) \quad \varrho^* = \inf\{J(f) : f \in \mathbf{F}\}.$$

DEFINITION 4.7. A stationary policy  $f$  is said to be *bias-optimal* if it attains the infimum in (39), i.e.,

$$J(f) = \varrho^* \quad \text{and} \quad \bar{h}_f(x) = \bar{h}(x) \quad \text{for all } x \in \mathbf{X}.$$

REMARK 4.8. (a) From (18), we have  $J_n(f, x, h) = J_n(f, x) + E_x^f h(x_n)$  for every  $f \in \mathbf{F}$ . Thus, under the hypotheses of Theorem 3.5, it follows from (21) and Remark 2.9(c) that if  $f \in \mathbf{F}$  is AC-optimal, then

$$D^u(f, \cdot) = D^l(f, \cdot) = \bar{h}_f(x),$$

which, together with Proposition 4.5(d), implies

$$\bar{h}(x) = \inf_{\mathbf{F}} D^u(f, x) \quad \text{for all } x \in \mathbf{X}.$$

(b) To summarize, we may combine Theorem 3.5 and Proposition 4.5 to see that, under Assumptions 2.4, 2.5, and 2.8, the implications in Diagram 1 below hold. Moreover, if we restrict ourselves to work only with stationary policies, it is shown in Theorem 4.12 that also Diagram 2 holds.

DIAGRAM 1. We have

$$\begin{array}{ccc} \text{strong OO} \Rightarrow \text{OO} \Rightarrow \text{OC-optim.} \Rightarrow \text{F-SAO} \Leftrightarrow \text{canonical} & & \\ \Downarrow & & \Updownarrow \\ \text{D-strong AO} & & \text{AC-optim.} \end{array}$$

DIAGRAM 2. In  $\mathbf{F}$ , we have

$$\text{OO} \Leftrightarrow \text{D-strong AO} \Leftrightarrow \text{OC-optim.} \Leftrightarrow \text{bias-optim.}$$

THEOREM 4.9. Under the hypotheses of Theorem 3.5 (i.e., Assumptions 2.4, 2.5 and 2.8), there exists a stationary policy  $f^*$  that is D-strong AO in  $\mathbf{F}$ , i.e.,

$$(41) \quad D^u(f^*, x) = \inf_{\mathbf{F}} D^u(f, x) =: D_0(x) < \infty \quad \text{for all } x \in \mathbf{X},$$

and it is the unique solution of the ACOE (22) for which

$$(42) \quad \int_{\mathbf{X}} D_0(y) Q_{f^*}(dy) = 0.$$

LEMMA 4.10. Under the hypotheses of Theorem 4.9, for each AC-optimal stationary policy  $f$  the following hold:

- (a)  $\text{OC}(f, \cdot) \leq \|D^u(f, \cdot)\|_{\mathbf{V}} b(1 - B)^{-1}$ , where  $b$  and  $B$  are the constants in (4), which are the same as in Remark 2.9(d);
- (b)  $D^u(f, \cdot) = L^u(\cdot) + \text{OC}^l(f, \cdot) = L^l(\cdot) + \text{OC}(f, \cdot)$ ;
- (c)  $\text{OC}^l(f, \cdot) = k_f = \liminf_{n \rightarrow \infty} [J_n(f, \cdot) - J_n^*(\cdot)]$   $Q_f$ -a.e. for some constant  $k_f \geq 0$ .

REMARK 4.11. Let  $(\varrho^*, h, f)$  be a canonical triplet with  $\int_{\mathbf{X}} h dQ_f = 0$ . An important problem related to Lemma 4.10 is the convergence of the *value iteration algorithm*. This is usually established by showing the convergence to a constant of the so-called *error functions*

$$(43) \quad e_n(x) := n\varrho^* + h(x) - J_n^*(x), \quad n \in \mathbb{N}_0, x \in \mathbf{X},$$

which, by (21), can also be written as

$$(44) \quad e_n(x) = J_n^*(x, h) - J_n^*(x) = J_n(f, x) - J_n^*(x) + E_x^f h(x_n).$$

The relation with Lemma 4.10 is that, since  $\lim_{n \rightarrow \infty} E_x^f h(x_n) = \int_{\mathbf{X}} h dQ_f = 0$ , we have

$$\lim_{n \rightarrow \infty} e_n(x) = k_f \quad \text{for all } x \in \mathbf{X}$$

if and only if

$$\text{OC}(f, x) = \text{OC}^l(f, x) = \lim_{n \rightarrow \infty} [J_n(f, x) - J_n^*(x)] = k_f \quad \text{for all } x \in \mathbf{X}.$$

Moreover, any of these conditions combined with Lemma 4.10(b) implies the following facts:

- (i)  $L(x) := L^u(x) = L^l(x)$  for all  $x \in \mathbf{X}$ ;
- (ii)  $D^u(f, \cdot) = L(\cdot) + k_f$ .

Finally, we have the theorem announced in Remark 4.8(b) [see Diagram 2].

THEOREM 4.12. *Under the assumptions of Theorem 4.9 the following statements are equivalent:*

- (i)  $f$  is  $D$ -strong AO in  $\mathbf{F}$ ;
- (ii)  $f$  is bias-optimal;
- (iii)  $f$  is OO in  $\mathbf{F}$ ;
- (iv)  $f$  is OC-optimal in  $\mathbf{F}$  and  $\text{OC}(f, x) < \infty$  for all  $x \in \mathbf{X}$ .

Consequently, from Theorem 4.9, there exists a stationary policy  $f$  that satisfies conditions (i)–(iv).

REMARK 4.13. Brown (1965) provides a very nice example showing that the results in Theorems 4.9 and 4.12 *cannot* be extended to the class  $\mathbf{\Pi}$  of *all* policies without additional hypotheses to Assumptions 2.4, 2.5 and 2.8. In fact, in his example there exists a stationary policy which is *Blackwell* optimal but it is not overtaking optimal in the class of *all* policies. We are grateful to Prof. A. S. Nowak for bringing Brown's example to our attention.

To close this section we present a slightly modified example by Nowak (1992), showing that in general, i.e., without the appropriate assumptions, some of the implications in Diagrams 1 and 2 may not hold. Specifically, we show the existence of a stationary policy  $f^*$  which is F-SAO and canonical,

but is neither OO nor OC-optimal. Moreover, there is another stationary policy  $f$  which is AC-optimal and F-SAO but is not canonical. To make the computations in the example we use the following two well-known facts from elementary stochastic Dynamic Programming:

(i) For any stationary policy  $f$  the  $n$ -stage expected total cost [see (10)] can be computed recursively as

$$(45) \quad J_n(f, x) = C(x, f) + \int_{\mathbf{X}} J_{n-1}(f, y) Q(dy | x, f) \quad \text{for all } x \in \mathbf{X}, n \in \mathbb{N};$$

recall that  $J_0(\pi, \cdot) := 0$  for any policy  $\pi$ . Similarly,

(ii) the  $n$ -stage value function [see (16)] satisfies

$$(46) \quad \begin{aligned} J_n^*(x) &= \min_{a \in \mathbf{A}(x)} \left[ C(x, a) + \int_{\mathbf{X}} J_{n-1}^*(y) Q(dy | x, a) \right] \\ &= (\mathbf{T}J_{n-1}^*)(x) \quad \text{for all } x \in \mathbf{X}, n \in \mathbb{N}, \end{aligned}$$

where  $\mathbf{T}$  is the operator in (1).

EXAMPLE 4.14. Consider a MCP with state space  $\mathbf{X} = \mathbb{N}_0$ , and control (or action) sets  $\mathbf{A}(\cdot) = \mathbf{A} = \{1, 2\}$ . The state  $x = 0$  is absorbing with zero cost, i.e. [writing  $Q(\{y\} | x, a)$  as  $Q(y | x, a)$ ],

$$(47) \quad C(0, \cdot) = 0 \quad \text{and} \quad Q(0 | 0, \cdot) = 1.$$

On the other hand, if  $x \geq 1$ , then

$$(48) \quad C(x, 1) = 1/x - 1, \quad Q(0 | x, 1) = 1,$$

and

$$(49) \quad C(x, 2) = 0, \quad Q(x+1 | x, 2) = 1.$$

From (47),  $J_n(\pi, 0) = 0$  for every policy  $\pi$  and  $n \in \mathbb{N}_0$ , so that  $J_n^*(0) = 0$  for all  $n \in \mathbb{N}_0$ . Moreover, since  $J_0^*(x) := 0$  for all  $x \in \mathbf{X}$ , we may use (46) to obtain

$$J_n^*(x) = (x + n - 1)^{-1} - 1 \quad \text{for all } x \geq 1, n \geq 1.$$

Similarly, from (45), the stationary policies  $f^*(\cdot) \equiv 2$  and  $f(\cdot) \equiv 1$  satisfy

$$J_n(f^*, x) = 0 \quad \text{for all } n \in \mathbb{N}_0, x \in \mathbf{X},$$

and

$$J_n(f, x) = 1/x - 1 \quad \text{for all } n \in \mathbb{N}, x \geq 1 \quad \text{and} \quad J_n(f, 0) = 0 \quad \text{for all } n \in \mathbb{N}_0.$$

In particular, from (26), the corresponding opportunity costs are

$$\text{OC}(f^*, x) = 0 \quad \text{for all } x \in \mathbf{X},$$

and

$$\text{OC}(f, x) = 1/x - 1 \quad \text{for all } x \geq 1, \quad \text{OC}(f, 0) = 0.$$

Now, straightforward calculations show  $f^*$  is a *canonical* policy since  $(\varrho^*, h, f^*)$ , with  $\varrho^* = 0$ ,  $h(0) = 0$  and  $h(x) = -1$  for  $x \geq 1$ , is a canonical triplet. Further, as  $h(\cdot)$  is a bounded function (hence it is in  $\mathbf{L}_V^\infty$ ),  $f^*$  is also AC-optimal and F-SAO. However,  $f^*$  is *not* OO because

$$\liminf_{n \rightarrow \infty} [J_n(f, x) - J_n(f^*, x)] = 1/x - 1 < 0 \quad \text{for } x > 1,$$

and, therefore,  $f^*$  does not satisfy (25). In addition,  $f^*$  is *not* OC-optimal since

$$\text{OC}^*(x) \leq \text{OC}(f, x) < \text{OC}(f^*, x) \quad \text{for all } x > 1.$$

Finally, as  $f$  has a finite opportunity cost, it follows [by Proposition 4.5(c) and Remark 3.3(b)] that  $f$  is F-SAO and AC-optimal; however,  $f$  is *not* canonical.

## 5. Proofs

*Proof of Proposition 2.6.* (a) Let  $u$  be an arbitrary function in  $\mathbf{L}_V^\infty$ , and let  $m := \|u\|_V$ . Then

$$u_V(x) := u(x) + m\mathbf{V}(x)$$

is a *nonnegative* function in  $\mathbf{L}_V^\infty$ . Since every nonnegative measurable function on  $\mathbf{X}$  is the limit of an increasing sequence of measurable bounded functions, a straightforward argument using Assumption 2.4(c) and the Monotone Convergence Theorem shows that, for every  $x \in \mathbf{X}$ ,

$$\int_{\mathbf{X}} u_V(y) Q(dy | x, \cdot) \text{ is l.s.c. on } \mathbf{A}(x).$$

Combining this fact with Assumptions 2.4(b) and (d) we see that

$$\begin{aligned} C(x, \cdot) + \int_{\mathbf{X}} u(y) Q(dy | x, \cdot) \\ = C(x, \cdot) + \int_{\mathbf{X}} u_V(y) Q(dy | x, \cdot) - m \int_{\mathbf{X}} \mathbf{V}(y) Q(dy | x, \cdot) \end{aligned}$$

is l.s.c. on  $\mathbf{A}(x)$  for every  $x \in \mathbf{X}$ . The last fact together with Assumption 2.4(a) yields—by a well-known Measurable Selection Theorem [see e.g. Himmelberg *et al.* (1976), Rieder (1978)]—the existence of a decision function  $f \in \mathbf{F}$  that satisfies (3).

(b) Now, since

$$\left| \int_{\mathbf{X}} u(y) Q(dy | x, a) \right| \leq m \int_{\mathbf{X}} \mathbf{V}(y) Q(dy | x, a),$$

we may use Assumptions 2.4(e) and 2.5(a) to conclude that  $\mathbf{T}u$  is in  $\mathbf{L}_V^\infty$ .

(c) Finally, Assumptions 2.4(a) and (d) yield—using the Measurable Selection Theorem again—that there is  $g \in \mathbf{F}$  such that



$$\sup_{a \in \mathbf{A}(x)} \int_{\mathbf{X}} \mathbf{V}(y) Q(dy | x, a) = \int_{\mathbf{X}} \mathbf{V}(y) Q(dy | x, g)$$

for every  $x \in \mathbf{X}$ , which combined with Assumption 2.5(a) implies (4) with constants  $B = B_g$  and  $b = b_g$ , and “petite set”  $C = C_g$ . ■

To prove Theorem 3.5 we shall use the following result whose proof is omitted here because it is essentially the same as that of Theorem 2.6 in Gordienko and Hernández-Lerma (1995a).

LEMMA 5.1. *Under the hypotheses of the Theorem 3.5 there is a real number  $\varrho^*$ , a function  $h$  in  $\mathbf{L}_{\mathbf{V}}^{\infty}$ , and a stationary policy  $f$  that satisfy the Average Cost Optimality Inequality (ACOI)*

$$(50) \quad \begin{aligned} \varrho^* + h(x) &\geq \min_{a \in \mathbf{A}(x)} \left[ C(x, a) + \int_{\mathbf{X}} h(y) Q(dy | x, a) \right] \\ &= C(x, f) + \int_{\mathbf{X}} h(y) Q(dy | x, f) \quad \text{for all } x \in \mathbf{X}. \end{aligned}$$

Moreover,  $f$  is AC-optimal and  $\varrho^*$  is the AC-value function, i.e. [as in (23)],

$$(51) \quad J(f, x) = J^*(x) = \varrho^* \quad \text{for all } x \in \mathbf{X}.$$

We next proceed to prove Theorem 3.5.

*Proof of Theorem 3.5.* (a) Let  $\varrho^*$ ,  $h$  and  $f$  be as in Lemma 5.1; we shall prove that equality holds in (50), for which we use the Poisson equation (6).

First note that integration of both sides of (6) with respect to the invariant probability measure  $Q_f$  and the use of (51) yield

$$(52) \quad J(f) = \int_{\mathbf{X}} C(y, f) Q_f(dy) = \varrho^*.$$

Hence, in lieu of (6), the Poisson equation when using the policy  $f$  can be written as

$$(53) \quad \varrho^* + h_f(x) = C(x, f) + \int_{\mathbf{X}} h_f(y) Q(dy | x, f), \quad x \in \mathbf{X},$$

for some function  $h_f$  in  $\mathbf{L}_{\mathbf{V}}^{\infty}$ . Thus, from (50) and (53), the function  $u$  in  $\mathbf{L}_{\mathbf{V}}^{\infty}$  defined as

$$u(x) := h(x) - h_f(x), \quad x \in \mathbf{X},$$

satisfies

$$(54) \quad u(x) \geq \int_{\mathbf{X}} u(y) Q(dy | x, f) \quad x \in \mathbf{X}.$$

Iterating this inequality (or noting that  $\{u(x_n)\}$  is a  $P_x^f$ -super-martingale) yields

$$u(x) \geq E_x^f u(x_n) \quad \text{for all } x \in \mathbf{X}, n \in \mathbb{N}_0,$$

and letting  $n \rightarrow \infty$  we obtain [by (5) or Remark 2.9(c)]

$$(55) \quad u(x) \geq \int_{\mathbf{X}} u(y) Q_f(dy) \quad \text{for all } x \in \mathbf{X}.$$

Therefore, defining  $k := \inf_{\mathbf{X}} u(x)$ , we get  $k \geq \int_{\mathbf{X}} u(y) Q_f(dy) \geq k$ , which implies that  $u(\cdot) = k$  ( $Q_f$ -a.e.). In other words [see Remark 2.9(a)],  $h$  and  $h_f$  differ only by the constant  $k$ , so that (53) can be written as

$$(56) \quad \varrho^* + h(x) = C(x, f) + \int_{\mathbf{X}} h(y) Q(dy | x, f) \quad \text{for all } x \in \mathbf{X};$$

that is, equality holds in (50). This proves that  $(\varrho^*, h, f)$  satisfies the ACOE (22), and the other statements in part (a) of the theorem follow from Lemma 5.1. [Concerning the equivalence of (21) and (22), see Remark 3.6(a).]

(b) We prove this part showing that the implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i) hold.

(i) $\Rightarrow$ (ii). This proof is very similar to that of part (a). Namely, if  $f \in \mathbf{F}$  is AC-optimal, then we have (52) and (53). Now, from (53) and Proposition 2.6(a), there is a stationary policy  $f'$  such that

$$(57) \quad \begin{aligned} \varrho^* + h_f(x) &\geq \min_{a \in \mathbf{A}(x)} \left[ C(x, a) + \int_{\mathbf{X}} h_f(y) Q(dy | x, a) \right], \\ &= C(x, f') + \int_{\mathbf{X}} h_f(y) Q(dy | x, f') \quad \text{for all } x \in \mathbf{X}. \end{aligned}$$

Hence, using Lemma 5.1, we find that  $f'$  is AC-optimal, that is, it satisfies (51). Finally, write down the Poisson equation for  $f'$  and compare it with (57) [see (54), (55)] to conclude that  $h_f$  and  $h_{f'}$  are equal, except perhaps for an additive constant. Thus  $(\varrho^*, h, f)$ , with  $h = h_f$ , satisfies the ACOE.

(ii) $\Rightarrow$ (iii). Let  $f \in \mathbf{F}$  be a canonical policy. Thus, from (21), for all  $x \in \mathbf{X}$  and  $n \in \mathbb{N}_0$ ,

$$(58) \quad J_n(f, x) + E_x^f h(x_n) = \inf_{\Pi} [J_n(\pi, x) + E_x^\pi h(x_n)] \leq J_n^*(x) + \sup_{\Pi} E_x^\pi h(x_n).$$

Therefore

$$0 \leq J_n(f, x) - J_n^*(x) \leq \sup_{\Pi} E_x^\pi h(x_n) - E_x^f h(x_n),$$

and upon multiplying by  $1/n$  and letting  $n \rightarrow \infty$ , Remark 2.9(d) yields

$$\lim_{n \rightarrow \infty} \frac{1}{n} [J_n(f, x) - J_n^*(x)] = 0 \quad \text{for all } x \in \mathbf{X}.$$

That is,  $f$  is F-SAO. (Observe that we also obtain  $\lim_{n \rightarrow \infty} \frac{1}{n} J_n^*(x) = \varrho^*$  for all  $x \in \mathbf{X}$ .)

(iii) $\Rightarrow$ (i) follows from Remark 3.3(b). ■

REMARK 5.2. Let  $\mathbf{F}^*$  be the set of all AC-optimal stationary policies. Then, by Theorem 3.5(b), and letting  $\mathbf{T}$  be as in (1), we may write  $\mathbf{F}^*$  as

$$(59) \quad \mathbf{F}^* = \{f \in \mathbf{F} : f \text{ is canonical [i.e., (21) holds]}\} \\ = \left\{ f \in \mathbf{F} : C(x, f) + \int_{\mathbf{X}} h(y) Q(dy | x, f) = \mathbf{T}h(x) \text{ for all } x \in \mathbf{X} \right\}.$$

Observe also that if  $f$  is in  $\mathbf{F}^*$ , then—by (21) and (23)—we may rewrite (29) and (30) as

$$(60) \quad D^u(f, x) = D^l(f, x) = \lim_{n \rightarrow \infty} [J_n(f, x) - n\varrho^*] \\ = \lim_{n \rightarrow \infty} [h(x) - E_x^n h(x_n)] =: D(f, x) \quad \text{for all } x \in \mathbf{X},$$

where, by (5),

$$(61) \quad D(f, x) = h(x) - \int_{\mathbf{X}} h(y) Q_f(dy) \quad \text{for all } x \in \mathbf{X}.$$

Finally, it is interesting to note that the following proof of existence of a D-strong AO policy in  $\mathbf{F}$  is *exactly the same* as Nowak's (1992) proof that there exists an OO policy (in  $\mathbf{F}$ )! This immediately suggests of course the equivalence, in Theorem 4.12, of D-strong average optimality and overtaking optimality in  $\mathbf{F}$ .

*Proof of Theorem 4.9.* If  $f \in \mathbf{F}$  is a stationary policy which is *not* AC-optimal (i.e.,  $J(f) > \varrho^*$ , or equivalently,  $f \notin \mathbf{F}^*$ ), then Proposition 4.5(d) yields  $D^u(f, \cdot) = \infty$ . Hence, to minimize  $D^u(\cdot, x)$  over  $\mathbf{F}$  we may restrict ourselves to look for a minimizer in the set  $\mathbf{F}^*$  of AC-optimal policies. Thus, from (59)–(61) we wish to find  $f^* \in \mathbf{F}^*$  such that

$$D(f^*, x) = \inf_{\mathbf{F}^*} D(f, x) = h(x) + \inf_{\mathbf{F}^*} \left[ - \int_{\mathbf{X}} h(y) Q_f(dy) \right] \quad \text{for all } x \in \mathbf{X}.$$

In other words [as in Nowak (1992)], the problem of minimizing  $D(\cdot, x)$  over  $\mathbf{F}^*$  for all  $x \in \mathbf{X}$  is equivalent to looking for an AC-optimal policy in a *new Markov control model*  $(\mathbf{X}, \mathbf{A}, \{\mathbf{A}^*(x) : x \in \mathbf{X}\}, Q, \widehat{C})$ , where

$$\mathbf{A}^*(x) := \left\{ a \in \mathbf{A}(x) : C(x, a) + \int_{\mathbf{X}} h(y) Q(dy | x, a) = \mathbf{T}h(x) \right\}, \quad x \in \mathbf{X}, \\ \widehat{C}(x, a) := -h(x), \quad a \in \mathbf{A}^*(x), \quad x \in \mathbf{X},$$

and  $\mathbf{X}$ ,  $\mathbf{A}$  and  $Q$  are the same as in the original Markov control model. Moreover, under the hypotheses of Theorem 4.9, the new control model obviously satisfies the assumptions of Theorem 3.5 with  $\mathbf{A}(\cdot)$  and  $C$  replaced by  $\mathbf{A}^*(\cdot)$  and  $\widehat{C}$ , respectively. Therefore, there exists an AC-optimal stationary

policy  $f^* \in \mathbf{F}^*$  for the new model, i.e.,

$$(62) \quad D(f^*, x) = h(x) - \int_{\mathbf{X}} h(y) Q_{f^*}(dy) = \inf_{\mathbf{F}^*} D(f, x) \quad \text{for all } x \in \mathbf{X}.$$

This shows that  $f^*$  is D-strong AO in  $\mathbf{F}^*$  [see (59) and (60)]. But, clearly,  $f^*$  is also D-strong AO in the set  $\mathbf{F}$  of *all* stationary policies since (as already noted) if  $f$  is in  $\mathbf{F} \setminus \mathbf{F}^*$ , then  $D^u(f, \cdot) = \infty$ . Thus, we have

$$(63) \quad D^u(f^*, x) = \inf_{\mathbf{F}} D^u(f, x) \quad \text{for all } x \in \mathbf{X}.$$

Finally, the last statement in Theorem 4.9 follows from (42) and Remark 3.6(d). ■

*Proof of Lemma 4.10.* Let  $f$  be an AC-optimal stationary policy. Then, by Theorem 3.5(b), it enters into some canonical triplet  $(\varrho^*, f, h)$ , where the function  $h$  can be chosen satisfying  $\int_{\mathbf{X}} h(y) Q_f(dy) = 0$ ; that is,  $h(\cdot) = \bar{h}_f(\cdot) = D^u(f, \cdot)$ ; see Remark 4.8(a).

(a) From (58), we have

$$0 \leq J_n(f, x) - J_n^*(x) \leq \sup_{\Pi} E_x^\pi h(x_n) - E_x^f h(x_n),$$

which, from Remark 2.9(c)–(d), implies that

$$\text{OC}(f, x) \leq \|h\|_{\mathbf{V}} b(1 - B)^{-1} \quad \text{for all } x \in \mathbf{X}.$$

(b) Now consider the error functions  $e_n$  in (43) and observe that [by (33), (34)]

$$\limsup_{n \rightarrow \infty} e_n(x) = h(x) - L^l(x) = \text{OC}(f, x) \quad \text{for all } x \in \mathbf{X},$$

and

$$\liminf_{n \rightarrow \infty} e_n(\cdot) = h(x) - L^u(x) = \text{OC}^l(f, x) \quad \text{for all } x \in \mathbf{X}.$$

Combining these facts with (60), which yields  $D^u(f, \cdot) = h(\cdot)$ , we obtain (b).

(c) As is well known [see Lemma 5.6.4 in Hernández-Lerma and Lasserre (1996), or Lemma 5.4 in Montes-de-Oca and Hernández-Lerma (1996)],

$$e_{n+k}(x) \geq \int_{\mathbf{X}} e_n(y) Q^k(dy | x, f) \quad \text{for all } n, k \in \mathbb{N}_0, x \in \mathbf{X}.$$

Thus, taking  $\liminf$  as  $k \rightarrow \infty$ , we obtain

$$(64) \quad \text{OC}^l(f, x) \geq \int_{\mathbf{X}} e_n(y) Q_f(dy) \quad \text{for all } n \in \mathbb{N}_0, x \in \mathbf{X},$$

which, by Fatou's Lemma, implies

$$\text{OC}^l(f, x) \geq \int_{\mathbf{X}} \text{OC}^l(f, y) Q_f(dy) \quad \text{for all } x \in \mathbf{X}.$$

[Note that Fatou's Lemma is indeed applicable since, by (44),  $e_n(x) \geq E_x^f h(x_n)$  and, therefore, by Remark 2.9(d),  $e_n$  is minorized by a  $Q_f$ -integrable function.] Hence

$$\text{OC}^l(f, \cdot) = \inf_{y \in \mathbf{X}} \text{OC}^l(f, y) =: k_f \quad Q_f\text{-a.e.},$$

which yields the first equality in (c). To obtain the second equality use (64) again, together with (44) and Fatou's Lemma, to obtain, for all  $x \in \mathbf{X}$ ,

$$\begin{aligned} \text{OC}^l(f, x) &\geq \liminf_{n \rightarrow \infty} \int_{\mathbf{X}} e_n(y) Q_f(dy) \geq \int_{\mathbf{X}} \liminf_{n \rightarrow \infty} e_n(y) Q_f(dy) \\ &= \int_{\mathbf{X}} \liminf_{n \rightarrow \infty} [J_n(f, y) - J_n^*(y)] Q_f(dy) = \int_{\mathbf{X}} \text{OC}^l(f, y) Q_f(dy). \end{aligned}$$

Therefore, the second equality in (c) follows from the first one. ■

*Proof of Theorem 4.12.* The equivalence of (i) and (ii) is an immediate consequence of Remark 4.8(a), while that of (i) and (iv) comes from Lemma 4.10(b) and Proposition 4.5(d). Indeed, from Lemma 4.10(b), we have

$$\inf_{\mathbf{F}^*} D^u(f, x) = L^l(x) + \inf_{\mathbf{F}^*} \text{OC}^u(f, x) \quad \text{for all } x \in \mathbf{X},$$

and using Proposition 4.5(d) we conclude that

$$\inf_{\mathbf{F}} D^u(f, x) = L^l(x) + \inf_{\mathbf{F}} \text{OC}^u(f, x) \quad \text{for all } x \in \mathbf{X}.$$

To obtain the equivalence of (i) and (iii), note that—as in Proposition 4.5(a)—overtaking optimality in  $\mathbf{F}$  implies D-strong AO in  $\mathbf{F}$ ; then it remains to prove that (i)  $\Rightarrow$  (iii). To do this, note that the following facts hold:

(a) for all  $f, g \in \mathbf{F}^*$ ,

$$\lim_n [J_n(f, x) - J_n(g, x)] = D^u(f, x) - D^u(g, x) \quad \text{for all } x \in \mathbf{X}.$$

(b) Consider policies  $f \in \mathbf{F}^*$  and  $g \in \mathbf{F} \setminus \mathbf{F}^*$ . Then

$$\limsup_{n \rightarrow \infty} [J_n(f, x) - J_n(g, x)] \leq D^u(f, x) - D^l(g, x) \leq 0 \quad \text{for all } x \in \mathbf{X},$$

since  $D^u(g, \cdot) = D^l(g, \cdot) = \infty$ .

Thus, from (a)–(b), we conclude that (i)  $\Rightarrow$  (iii). ■

**6. An example.** We now discuss an example from inventory theory to illustrate how to verify our assumptions. For further details see Vega-Amaya (1996).

Consider a discrete time inventory system in which the stock level  $x_t$  evolves in  $\mathbf{X} := [0, \infty)$  according to

$$(65) \quad x_{t+1} = \max(x_t + a_t - w_t, 0), \quad t \in \mathbb{N}_0; \quad x_0 = x.$$

Here  $a_t$  denotes the amount of product ordered (and immediately supplied) at the beginning of each period  $t \in \mathbb{N}_0$ , whereas  $w_t$  is the product's demand during that period. We suppose that the production variables  $\{a_t\}$  take values in the interval  $\mathbf{A} := [0, \theta]$  irrespective of the stock level, where  $\theta$  is a positive constant.

We suppose that the demand process  $\{w_t\}$  satisfies the following assumption:

ASSUMPTION 6.1. (a) The process  $\{w_t\}$  is formed by nonnegative independent identically distributed random variables with common cumulative distribution function denoted by  $\phi$ , and such that

- (a.1)  $\phi$  has a continuous bounded density function  $\mu$ ;
- (a.2)  $w^* := \int_0^\infty y \phi(dy) < \infty$ .

(b)  $\theta < w^*$ .

In what follows,  $E$  denotes the expectation with respect to  $\phi$ .

The one-step (net) cost function is given by

$$(66) \quad C(x, a) := ba + h_c(x + a) - sE \min(x + a, w_0), \quad (x, a) \in \mathbf{K} = \mathbf{X} \times \mathbf{A},$$

where  $b$ ,  $h_c$  and  $s$  are positive constants.

Now, we proceed to show that, for the inventory system (65)–(66), Assumption 6.1 implies Assumptions 2.4, 2.5 and 2.8.

*Verification of Assumption 2.4.* It is clear that Assumption 2.4(a)–(b) holds; Assumption 2.4(c) follows from (a.1) and noting that

$$\int_{\mathbf{X}} u(y) Q(dy | x, a) = Eu[(x + a - w_0)^+],$$

for all  $(x, a) \in \mathbf{K}$ , and any measurable bounded  $u$  on  $\mathbf{X}$ , where  $y^+ := \max(0, y)$ .

To obtain a “bounding” function  $\mathbf{V}$  satisfying Assumption 2.4(d)–(e), first note that, for the function  $\Psi(p) := E \exp[p(\theta - w_0)]$ ,  $p \geq 0$ , we have  $\Psi(0) = 1$  and  $\Psi'(0) = \theta - w^* < 0$ , which implies that there exists a constant  $r > 0$  such that

$$(67) \quad \alpha := E \exp[r(\theta - w_0)] < 1.$$

Now, define

$$(68) \quad \mathbf{V}(x) := \beta \exp[r(x + 2w^*)], \quad x \in \mathbf{X},$$

where  $\beta$  is a positive constant. Simple computations yield

$$(69) \quad \int_{\mathbf{X}} \mathbf{V}(y) Q(dy | x, a) = \mathbf{V}(x) \int_0^{x+a} \exp[r(a-y)] \phi(dy) + \mathbf{V}(0)[1 - \phi(x+a)]$$

for all  $(x, a) \in \mathbf{K}$ , which implies Assumption 2.4(d).

On the other hand, direct computations yield that  $\sup_{a \in \mathbf{A}} |C(x, a)| \leq s(x + 2w^*)$  for all  $x \in \mathbf{X}$ . Thus, since one can choose the constant  $\beta$  large enough such that  $\mathbf{V}(x) \geq \max\{1, s(x + 2w^*)\}$  for all  $x \in \mathbf{X}$ , we see that Assumption 2.4(e) holds.

*Verification of Assumption 2.5.* First we shall show that Assumption 2.5(b) is satisfied. To do this, define

$$(70) \quad \psi(A) := \mathbf{I}_A(0), \quad A \in \mathcal{B}_{\mathbf{X}}, \quad \text{and} \quad S(x) := 1 - \phi(x + \theta), \quad x \in \mathbf{X}.$$

Using this notation, we have

$$(71) \quad Q(\cdot | x, a) \geq \psi(\cdot)S(x) \quad \text{for all } (x, a) \in \mathbf{K},$$

which implies that, for each  $f \in \mathbf{F}$ , the Markov processes  $\{x_t\}$  is  $\psi$ -irreducible and aperiodic [Nummelin (1984), Remark 2.1 and Example 2.5]; hence, Assumption 2.5(b) holds.

To verify Assumption 2.5(a), from (67)–(69), note that

$$\int_{\mathbf{X}} \mathbf{V}(y) Q(dy | x, a) \leq \alpha \mathbf{V}(x) + \mathbf{V}(0) \quad \text{for all } (x, a) \in \mathbf{K},$$

which implies

$$\int_{\mathbf{X}} \mathbf{V}(y) Q(dy | x, a) \leq B \mathbf{V}(x) + \mathbf{V}(0) \mathbf{I}_K(x) \quad \text{for all } (x, a) \in \mathbf{K},$$

where  $B := \frac{1}{2}(1 - \alpha) < 1$  and  $K := \{y \in \mathbf{X} : \mathbf{V}(y) \leq (1 - \alpha)^{-1} \mathbf{V}(0)\}$ .

Since  $S(\cdot)$  is a continuous function, inequality (71) implies that every compact subset of  $\mathbf{X}$  is a petite set for each  $f \in \mathbf{F}$  [Meyn and Tweedie (1993), Proposition 6.2.4 and Theorem 6.2.5, p.134; using the terminology of Meyn and Tweedie, (71) implies that, for each  $f \in \mathbf{F}$ , the process  $\{x_t\}$  is a *T-chain*]. Hence, since  $K$  is a compact subset of  $\mathbf{X}$ , it is a petite set for each policy  $f \in \mathbf{F}$ . Therefore, Assumption 2.5(a) holds.

*Verification of Assumption 2.8.* To check this part we shall use a result from Gordienko and Hernández-Lerma (1995a), for which we introduce the following notation: let  $\nu$  be a finite signed measure on  $\mathbf{X}$  and define the norm

$$\|\nu\|_{\mathbf{V}}^* := \int_{\mathbf{X}} \mathbf{V}(y) |\nu|(dy),$$

where  $\mathbf{V}(\cdot)$  is the function (68) and  $|\nu|$  is the total variation of the signed measure  $\nu$ . Moreover, define

$$\nu_f^n(\cdot | x) := Q^n(\cdot | x, f) - Q_f(\cdot), \quad x \in \mathbf{X}, \quad f \in \mathbf{F}, \quad n \in \mathbb{N},$$

and

$$S_f(x) := 1 - \phi(x + f(x)), \quad x \in \mathbf{X}, \quad f \in \mathbf{F}.$$

Observe that the condition

$$(72) \quad \sup_{f \in \mathbf{F}} \|\nu_f^n(\cdot | x)\|_{\mathbf{V}}^* \leq \mathbf{V}(x)M\gamma^n, \quad \text{for all } x \in \mathbf{X}, n \in \mathbb{N},$$

implies that Assumption 2.8 holds.

Now, note that for each  $x \in \mathbf{X}$  and  $f \in \mathbf{F}$ , the following conditions hold:

- (i)  $Q(\cdot | x, f) \geq S_f(x)\psi(\cdot)$ ;
- (ii)  $\int_{\mathbf{X}} \mathbf{V}(y) Q(dy | x, f) \leq \alpha \mathbf{V}(x) + \|\psi\|_{\mathbf{V}}^* S_f(x)$  [from (69)];
- (iii)  $\inf_{f \in \mathbf{F}} \int_{\mathbf{X}} S_f(y) \psi(dy) \geq \int_{\mathbf{X}} S(y) \psi(dy) = S(\theta) > 0$ .

Gordienko and Hernández-Lerma [(1995a), Lemma 3.4] show that properties (i)–(iii) imply (72); hence, Assumption 2.8 is satisfied.

**7. Concluding remarks and open problems.** In the previous sections we have presented an analysis of several infinite-horizon, undiscounted optimality criteria for a class of discrete-time Markov control processes (MCPs) on *Borel* spaces. Some of our hypotheses—especially Assumption 2.5—may seem to be quite restrictive, but the fact is that weakening these hypotheses does not look very promising. Indeed, a glance at the literature [e.g., Dutta (1991), Fernández-Gaucheraud *et al.* (1994) and Nowak (1992) for discrete-time problems, or Carlson *et al.* (1991) and Leizarowitz (1988) in the continuous-time case] shows that implicitly or explicitly one needs conditions ensuring results such as the exponential convergence in (5) and/or the existence of solutions to the Poisson equation (6).

On the other hand, once we have the last (strong) conditions—say (5) and (6)—one would definitely expect to solve other interesting open problems. For example, does the *value iteration* algorithm converge? (See Lemma 4.10 and Remark 4.11.) Can one (say, “similarly”) obtain the convergence of the *policy iteration* (PI) algorithm? This is a standard technique for MCPs with a *denumerable* (mainly finite) state space, but, to our knowledge, there are no results whatsoever for the *Borel* space case, except for recent work of Meyn (1995) and Hernández-Lerma and Lasserre (1997). [For *discounted cost* problems see, e.g., Hernández-Lerma and Lasserre (1996) or Hernández-Lerma and Muñoz de Ozak (1992).] Similarly, (5) and (6) also yield the Strong Law of Large Numbers, the Central Limit Theorem (CLT) and the Law of the Iterated Logarithm (LIL) for both discrete and continuous time Markov processes [Glynn and Meyn (1996), Meyn and Tweedie (1993)]. Hence one would “expect” to obtain, for discrete and continuous MCPs, “pathwise” (or sample path) versions of Theorems 3.5 and 4.9 [cf. Leizarowitz (1988)], and/or the Borel-space version of the CLT and the LIL obtained, e.g., by Mandl and Lausmanová (1991) for finite-state MCPs.

In conclusion, it is apparent that the setting and techniques developed in this paper may be useful to solve other MCP-related problems.



**Acknowledgments.** We wish to thank Professor Andrzej S. Nowak for important comments on a previous version of this paper.

### References

- A. Arapostathis, V. S. Borkar, E. Fernández-Gaucherand, M. K. Ghosh and S. I. Marcus (1993), *Discrete-time controlled Markov processes with average cost criterion: a survey*, SIAM J. Control Optim. 31, 282–344.
- R. Bellman (1957), *A markovian decision process*, J. Math. Mech. 6, 679–684.
- D. P. Bertsekas and S. E. Shreve (1978), *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York.
- B. W. Brown (1965), *On the iterative method of dynamic programming on a finite space discrete time Markov process*, Ann. Math. Statist. 33, 719–726.
- D. A. Carlson, A. Haurie and A. Leizarowitz (1991), *Infinite Horizon Optimal Control: Deterministic and Stochastic Systems*, Springer, New York.
- E. V. Denardo and U. G. Rothblum (1979), *Overtaking optimality for Markov decision chains*, Math. Oper. Res. 4, 144–152.
- P. K. Dutta (1991), *What do discounted optima converge to? A theory of discount rate asymptotics in economic models*, J. Econom. Theory 55, 64–94.
- E. B. Dynkin and A. A. Yushkevich (1979), *Controlled Markov Processes*, Springer, New York.
- A. Ephremides and S. Verdú (1989), *Control and optimization methods in communication network problems*, IEEE Trans. Automat. Control 34, 930–942.
- E. Fernández-Gaucherand, M. K. Ghosh and S. I. Marcus (1994), *Controlled Markov processes on the infinite planning horizon: weighted and overtaking criteria*, Z. Oper. Res. 39, 131–155.
- J. Flynn (1980), *On optimality criteria for dynamic programs with long finite horizons*, J. Math. Anal. Appl. 76, 202–208.
- D. Gale (1967), *On optimal development in a multi-sector economy*, Rev. Econom. Stud. 34, 1–19.
- P. W. Glynn and S. P. Meyn (1996), *A Lyapunov bound for solutions of Poisson's equation*, Ann. Probab. 24, 916–931.
- E. Gordienko and O. Hernández-Lerma (1995a), *Average cost Markov control processes with weighted norms: existence of canonical policies*, Appl. Math. (Warsaw) 23, 199–218.
- E. Gordienko and O. Hernández-Lerma (1995b), *Average cost Markov control processes with weighted norms: value iteration*, *ibid.*, 219–237.
- O. Hernández-Lerma (1989), *Adaptive Markov Control Processes*, Springer, New York.
- O. Hernández-Lerma, J. C. Hennet and J. B. Lasserre (1991), *Average cost Markov decision processes: optimality conditions*, J. Math. Anal. Appl. 158, 396–406.
- 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. 47, 125–154.
- O. Hernández-Lerma, R. Montes-de-Oca and R. Cavazos-Cadena (1991), *Recurrence conditions for Markov decision processes with Borel state space: a survey*, Ann. Oper. Res. 28, 29–46.

- O. Hernández-Lerma and M. Muñoz de Ozak (1992), *Discrete-time Markov control processes with discounted unbounded cost: optimality criteria*, *Kybernetika* (Prague) 28, 191–212.
- C. J. Himmelberg, T. Parthasarathy and F. S. Van Vleck (1976), *Optimal plans for dynamic programming problems*, *Math. Oper. Res.* 1, 390–394.
- A. Leizarowitz (1988), *Controlled diffusion processes on infinite horizon with the overtaking criterion*, *Appl. Math. Optim.* 17, 61–78.
- P. Mandl and M. Lausmanová (1991), *Two extensions of asymptotic methods in controlled Markov chains*, *Ann. Oper. Res.* 28, 67–79.
- S. P. Meyn (1995), *The policy improvement algorithm for Markov decision processes with general state space*, preprint, Coordinated Science Laboratory, Univ. of Illinois, Urbana, Ill.
- S. P. Meyn and R. L. Tweedie (1993), *Markov Chains and Stochastic Stability*, Springer, London.
- R. Montes-de-Oca and O. Hernández-Lerma (1996), *Value iteration in average cost Markov control processes on Borel spaces*, *Acta Appl. Math.* 42, 203–222.
- A. S. Nowak (1992), *Stationary overtaking optimal strategies in Markov decision processes with general state space*, preprint, Institute of Mathematics, Technical Univ. of Wroc/law.
- E. Nummelin (1984), *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge Univ. Press, Cambridge.
- S. Orey (1971), *Lecture Notes on Limit Theorems for Markov Chain Transition Probabilities*, Van Nostrand Reinhold, London.
- M. L. Puterman (1994), *Markov Decision Processes*, Wiley, New York.
- F. P. Ramsey (1928), *A mathematical theory of savings*, *Econom. J.* 38, 543–559.
- U. Rieder (1978), *Measurable selection theorems for optimization problems*, *Manuscripta Math.* 24, 115–131.
- P. J. Schweitzer (1985), *On undiscounted Markovian decision processes with compact action spaces*, *RAIRO Rech. Opér.* 19, 71–86.
- S. Stidham and R. Weber (1993), *A survey of Markov decision models for control of networks of queues*, *Queueing Systems Theory Appl.* 13, 291–314.
- O. Vega-Amaya (1996), *Overtaking optimality for a class of production-inventory systems*, preprint, Departamento de Matemáticas, Universidad de Sonora.
- A. F. Veinott, Jr. (1966), *On finding optimal policies in discrete dynamic programming with no discounting*, *Ann. Math. Statist.* 37, 1284–1294.
- C. C. von Weizsäcker (1965), *Existence of optimal programs of accumulation for an infinite horizon*, *Rev. Econom. Stud.* 32, 85–104.
- A. A. Yushkevich (1973), *On a class of strategies in general Markov decision models*, *Theory Probab. Appl.* 18, 777–779.

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

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

*Received on 6.3.1997;  
 revised version on 15.7.1997*