APPLICATIONES MATHEMATICAE
23,2 (1995), pp. 199–218

E. GORDIENKO and O. HERNÁNDEZ-LERMA (México)

# AVERAGE COST MARKOV CONTROL PROCESSES WITH WEIGHTED NORMS: EXISTENCE OF CANONICAL POLICIES

*Abstract*. This paper considers discrete-time Markov control processes on Borel spaces, with possibly unbounded costs, and the long run *average cost* (AC) criterion. Under appropriate hypotheses on weighted norms for the cost function and the transition law, the existence of solutions to the *average cost optimality inequality* and the *average cost optimality equation* are shown, which in turn yield the existence of AC-*optimal* and AC-*canonical* policies respectively.

**1. Introduction.** Among the several approaches to prove the existence of *average cost optimal* (hereafter abbreviated *AC-optimal*) policies for Markov control processes (MCPs) two of the most widely used are the so-called *vanishing discount* approach, and the one based on *strong ergodicity* assumptions. In the former, the idea is to impose conditions on an associated $\beta$-*discounted cost* problem in such a way that as $\beta \uparrow 1$ we obtain in the limit either the *average cost optimality inequality* (ACOI) or the *average cost optimality equation* (ACOE), each of which in turn yields an AC-optimal policy (see e.g. [1, 7, 8, 9, 16, 24]). On the other hand, imposing strong ergodicity assumptions usually allows one to obtain directly the ACOE; this approach, however, has been mainly used for MCPs with *bounded* cost functions [1, 3, 6, 10].

In this paper we combine the two approaches to obtain the ACOI and

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

*Key words and phrases*: (discrete-time) Markov control processes, long run average cost, discounted cost, weighted norms, average cost optimality inequality, average cost optimality equation.

This research was supported in part by the Consejo Nacional de Ciencia y Tecnología (CONACYT) Grant 1332-E9206. The research of the first author was also supported by CONACYT Professorship (Catedra Patrimonial Nivel II) No. 400261-2.

the ACOE for MCPs on Borel spaces with possibly unbounded (from above) cost functions. Namely, we impose ergodicity hypotheses under which the vanishing discount approach is applicable. The main difference between our paper and the previous literature is that, following Kartashov [12, 13], the ergodicity conditions we use are expressed in terms of *weighted* norms. This allows, in particular, "differential cost" functions in the ACOI or the ACOE (see $\phi^*$ in (2.8), (2.10)) which may be unbounded.

Our main hypotheses and results (Theorems 2.6 and 2.8) are presented in Section 2. Sections 3–6 contain the proofs of Theorems 2.6 and 2.8, and important corollaries of these results are stated in Section 7. Finally, Section 8—an appendix—summarizes some results for Harris-recurrent Markov chains, which are needed in the statement of our hypotheses.

## 2. The control model and main results

Remark 2.1. If $X$ is a *Borel space* (i.e., a Borel subset of a complete and separable metric space), its Borel $\sigma$-algebra is denoted by $\mathbb{B}_X$. Let $X$ and $Y$ be Borel spaces. A *stochastic kernel* [1, 2, 3, 6, 9] (or transition probability function) on $X$ given $Y$ is a function $P(B \mid y)$ such that $P(\cdot \mid y)$ is a probability measure on $\mathbb{B}_X$ for each fixed $y \in Y$, and $P(B \mid \cdot)$ is a measurable function on $Y$ for each fixed $B \in \mathbb{B}_X$.

Let $(X, A, Q, c)$ be a discrete-time Markov control model with state space $X$, action (or control) set $A$, transition law $Q$, and one-stage cost function $c$ satisfying the following conditions (cf. [1, 2, 3, 6, 9, 11]). $X$ and $A$ are both Borel spaces. For each $x \in X$ there is a nonempty Borel set $A(x)$ in $\mathbb{B}_A$ which represents the set of feasible actions in the state $x$. The set

$$\mathbb{K} := \{(x, a) : x \in X, a \in A(x)\} \tag{2.1}$$

is assumed to be a Borel subset of $X \times A$. The transition law $Q$ is a stochastic kernel on $X$ given $\mathbb{K}$ and, finally, the one-stage cost $c$ is a real-valued measurable function on $\mathbb{K}$.

The interpretation of $(X, A, Q, c)$ as representing a MCP is well known (see the references in the previous section).

Assumption 2.2. (a) The one-stage cost $c$ is nonnegative and $a \mapsto c(x, a)$ is lower semicontinuous (l.s.c.) on $A(x)$ for each $x \in X$; moreover, there exists a measurable function $v : X \to \mathbb{R}$ such that $\overline{v} := \inf_X v(x) > 0$,

$$\sup_{A(x)} c(x, a) \le v(x) \quad \forall x \in X, \tag{2.2a}$$

$$\int_X v(y)\, Q(dy \mid x, a) < \infty \quad \forall (x, a) \in \mathbb{K}, \tag{2.2b}$$

and

(2.2c) $$a \mapsto \int_X v(y)\, Q(dy \mid x, a) \text{ is continuous on } A(x) \text{ for every } x \in X;$$

(b) $A(x)$ is compact for each state $x$;

(c) the transition law $Q$ is *strongly continuous* on $A$, i.e., for each measurable and bounded function $u$ on $X$ and each state $x$, the map

$$a \mapsto \int_X u(y)\, Q(dy \mid x, a)$$

is continuous on $A(x)$.

Let $\Delta$ the class of all (possibly randomized and nonstationary) *control policies* [1, 2, 3, 6, 9, 11], and let $\Delta_0$ be the subclass of (deterministic) *stationary* policies. By a standard convention, we will identify $\Delta_0$ with the class of all measurable functions $f : X \to A$ such that $f(x) \in A(x)$ for all $x$ in $X$. ($\Delta_0$ is nonempty: see Example 2.6 in [21].)

Remark. Let $f \in \Delta_0$ be an arbitrary stationary policy. Then, when using $f$, the state process is a *Markov* chain with transition kernel $Q(\cdot \mid x, f(x))$, which will also be written as $Q_f(\cdot \mid x)$ or $Q(\cdot \mid x, f)$, i.e.,

(2.3a) $$Q(\cdot \mid x, f(x)) \equiv Q_f(\cdot \mid x) \equiv Q(\cdot \mid x, f), \quad x \in X.$$

We will also use the notation

(2.3b) $$c(x, f(x)) = c(x, f), \quad x \in X.$$

Let $P_x^\delta$ be the induced probability measure when using the policy $\delta \in \Delta$ given the initial state $x_0 = x$ (see e.g. Hinderer [11, p. 80] for the construction of $P_x^\delta$).

The corresponding expectation operator is denoted by $E_x^\delta$.

Let

(2.4) $$J_n(\delta, x) := E_x^\delta \Big[ \sum_{t=0}^{n-1} c(x_t, a_t) \Big], \quad n = 1, 2, \ldots,$$

be the expected *n-stage cost* when using the policy $\delta$, given the initial state $x_0 = x$, and let

(2.5) $$J(\delta, x) := \limsup_{n \to \infty} \frac{J_n(\delta, x)}{n}$$

be the corresponding long-run expected *average cost* (AC) per unit time.

A policy $\delta^*$ is said to be *AC-optimal* if $J(\delta^*, x) = J^*(x)$ for all $x \in X$, where

$$J^*(x) := \inf_{\Delta} J(\delta, x), \quad x \in X,$$

is the *optimal AC-function*. (In Remark 2.9 we introduce the stronger concept of *canonical* policy.)

The *main problem* we are concerned with is precisely to show the existence of AC-optimal (and canonical) policies. To do this we shall require several hypotheses, in addition to Assumption 2.2. In particular, in Assumptions 2.3 and 2.4 below, we use the notion of Harris recurrence (see Section 8) and the notation (2.3).

ASSUMPTION 2.3. For each stationary policy $f \in \Delta_0$ the (state) Markov process defined by the stochastic kernel $Q_f$ in (2.3a) is positive Harris-recurrent, i.e., it is Harris-recurrent and has an invariant probability measure $q_f$:

$$q_f(B) = \int_X Q_f(B \mid x)\, q_f(dx) \quad \forall B \in \mathbb{B}_X. \tag{2.6}$$

By Assumption 2.3 and Proposition 8.1(b), for each $f$ in $\Delta_0$ there exists a triplet $(n_f, \nu_f, h_f)$ consisting of an integer $n_f$, a probability measure $\nu_f$ on $X$, and a nonnegative measurable function $h_f \le 1$ on $X$ which satisfies the analogue of inequalities (i)–(iii) in Proposition 8.1. The following assumption requires that these inequalities are satisfied uniformly in $f \in \Delta_0$ with $n_f \equiv 1$, and requires the function $v$ in Assumption 2.2 to be analogous to a so-called $\alpha$-excessive function.

ASSUMPTION 2.4. There exist a probability measure $\nu$ on $X$ and a number $0 \le \alpha < 1$ for which the following holds: For each $f \in \Delta_0$ there is a nonnegative measurable function $h_f \le 1$ on $X$ such that, for every $x \in X$ and $B \in \mathbb{B}_X$:

(a) $Q_f(B \mid x) \ge h_f(x)\nu(B)$;
(b) $\int_X v(y)\, Q_f(dy \mid x) \le h_f(x)\|\nu\|_v + \alpha v(x)$, and

$$\|\nu\|_v := \int_X v(y)\, \nu(dy) < \infty; \tag{2.7}$$

(c) $\inf_{\Delta_0} \int_X h_f(x)\, \nu(dx) =: \gamma > 0$.

The notation in (2.7) will be used in the following more general sense.

DEFINITION 2.5. $M_v$ denotes the normed linear space consisting of all finite signed measures $\mu$ on X for which

$$\|\mu\|_v := \int_X v(x)\, |\mu|(dx) < \infty,$$

where $|\mu|$ stands for the total variation of $\mu$. Similarly, $\Phi_v$ denotes the normed linear space of all measurable fuctions $\phi : X \to \mathbb{R}$ with

$$\|\phi\|_v := \sup_X \frac{|\phi(x)|}{v(x)} < \infty.$$

We are now ready to state one of our main results.

THEOREM 2.6. *Suppose that Assumptions* 2.2–2.4 *hold. Then there exists a constant* $\varrho^* \geq 0$, *a function* $\phi^*$ *in* $\Phi_v$, *and a stationary policy* $f^*$ *such that*:

(a) *we have*

$$\varrho^* + \phi^*(x) \geq \min_{A(x)} \Big[c(x,a) + \int\limits_X \phi^*(y)\, Q(dy \mid x,a)\Big] \tag{2.8}$$

$$= c(x,f^*) + \int\limits_X \phi^*(y)\, Q(dy \mid x, f^*) \qquad \forall x \in X;$$

(b) $f^*$ *is AC-optimal and* $\varrho^*$ *is the optimal AC-function*, *i.e.*,

$$J(f^*,x) = J^*(x) = \varrho^* \qquad \forall x \in X. \tag{2.9}$$

The inequality in (2.8), which may be *strict* [9], is known as the *average cost optimality inequality* (ACOI). Our second main result, Theorem 2.8, gives conditions for *equality* to hold in (2.8), thus yielding the *average cost optimality equation* (ACOE)

$$\varrho^* + \phi^*(x) = \min_{A(x)} \Big[c(x,a) + \int\limits_X \phi^*(y)\, Q(dy \mid x,a)\Big]. \tag{2.10}$$

To state this result we need additional notation and assumptions.

Let $d_1$ and $d_2$ be the metrics on $X$ and $A$ respectively, and let $d$ be the metric on $\mathbb{K}$ (the set in (2.1)) defined as

$$d((x,a),(x',a')) := \max\{d_1(x,x'), d_2(a,a')\}, \tag{2.11}$$

for all $(x,a)$ and $(x',a')$ in $\mathbb{K}$. Moreover, let $\Psi$ be the class of nondecreasing functions $\psi : [0,\infty) \to [0,\infty)$ such that $\lim_{s\downarrow 0} \psi(s) = 0$.

In addition to Assumption 2.2 we shall suppose the following:

ASSUMPTION 2.7. (a) The compact-valued multifunction $x \mapsto A(x)$ is continuous with respect to the Hausdorff metric;

(b) for each $x$ in $X$, there exist functions $\psi_x^c$ and $\psi_x^Q$ in $\Psi$ such that, for all $a \in A(x)$ and $(x',a') \in \mathbb{K}$:

(i) $|c(k) - c(k')| \leq \psi_x^c[d(k,k')]$ and
(ii) $\|Q(\cdot \mid k) - Q(\cdot \mid k')\|_v \leq \psi_x^Q[d(k,k')]$,

where $k := (x,a)$ and $k' := (x',a')$.

An example of a MCP satisfying Assumptions 2.2–2.4 and 2.7 is presented in [5, Section 5].

THEOREM 2.8. *Suppose that the hypotheses of Theorem* 2.6, *as well as Assumption* 2.7 *are valid. Then there exists a constant* $\varrho^* \geq 0$, *a continuous function* $\phi^*$ *in* $\Phi_v$, *and a stationary policy* $f^*$ *such that*:

(a) $\varrho^*$ *and* $\phi^*$ *satify the ACOE* (2.10) *and*, *in addition*,

$$\varrho^* + \phi^*(x) = c(x, f^*) + \int\limits_X \phi^*(y)\, Q(dy \mid x, f^*) \qquad \forall x \in X; \tag{2.12}$$

(b) (2.9) *holds*, *i.e.*, $f^*$ *is AC-optimal and* $\varrho^*$ *is the optimal AC-function.*

R e m a r k 2.9. A triplet $(\varrho^*, \phi^*, f^*)$ satisfying (2.10) and (2.12) is called a *canonical triplet* [1, 3, 9, 17], and $f^* \in \Delta_0$ is said to be a *canonical policy*. According to Theorem 2.8(b), a canonical policy is AC-optimal, but the converse is *not* true in general; in other words (as shown in the above-given references), there are AC-optimal policies $f^*$ in $\Delta_0$ for which (2.10) and (2.12) do not hold.

The proofs of Theorems 2.6 and 2.8 are presented in Sections 3–6. Some important corollaries are given in Section 7.

We conclude this section with an elementary, but useful remark.

R e m a r k 2.10. The function $x \mapsto \int_X v(y)\, Q_f(dy \mid x)$ is in $\Phi_v$ for all $f \in \Delta_0$; in fact,

$$\sup_{\Delta_0} \sup_X \int\limits_X v(y)\, Q_f(dy \mid x)/v(x) \leq \|\nu\|_v/\overline{v} + \alpha. \tag{2.13}$$

Indeed, since $v(x) \geq \overline{v}$ for all $x \in X$ (see Assumption 2.2(a)) and $h_f(\cdot) \leq 1$, the inequality in Assumption 2.4(b) yields

$$\int\limits_X v(y)\, Q_f(dy \mid x) \leq (\|\nu\|_v/\overline{v} + \alpha) v(x),$$

which in turn yields (2.13). An inequality related to (2.13) is given in (3.2) below.

**3. Lemmas on ergodicity conditions.** Under Assumptions 2.3 and 2.4, Kartashov [14, Corollary 2] shows that the invariant probability measure $q_f$ has $\|q_f\|_v$ finite for every stationary policy $f$. In the following elementary lemma we show *inter alia* that $\|q_f\|_v$ is bounded above uniformly in $f \in \Delta_0$.

LEMMA 3.1. *Suppose that Assumptions* 2.3 *and* 2.4 *hold*, *and let* $v$ *and* $\overline{v}$ *be as in Assumption* 2.2(a). *Then for every stationary policy* $f \in \Delta_0$:

(a) $\gamma \leq \int_X h_f\, d\nu \leq 1$, *with* $\gamma$ *as in Assumption* 2.4(c);
(b) $\overline{v} \leq \|q_f\|_v \leq b_0$, *with* $b_0 := \|\nu\|_v/(1 - \alpha)$;
(c) $b_1 \leq \int_X h_f\, dq_f \leq 1$, *with* $b_1 := \overline{v}(1 - \alpha)/\|\nu\|_v = \overline{v}/b_0$.

P r o o f. (a) This follows from Assumption 2.4(c) and the fact that $h_f \le 1$.

(b), (c). The first inequality in (b) follows from the definition of $\overline{v}$, and the second inequality in (c) is due to $h_f \le 1$. On the other hand, from part (a), (2.6) and Assumption 2.4(b),

$$\begin{aligned}\|q_f\|_v &:= \int v(y) q_f(dy) = \int\int v(y)\, Q_f(dy \mid x)\, q_f(dx) \\ &\le \|\nu\|_v \int h_f\, dq_f + \alpha \|q_f\|_v.\end{aligned}$$

(Cf. Remark 2.10.) This implies, on the one hand, that

$$\|q_f\|_v \le \frac{\|\nu\|_v}{1-\alpha} = b_0 \qquad (\text{since } \textstyle\int h_f dq_f \le 1),$$

and, on the other hand,

$$\int h_f\, dq_f \ge \frac{(1-\alpha)\|q_f\|_v}{\|\nu\|_v} \ge b_1. \ \blacksquare$$

Let $f \in \Delta_0$ be an arbitrary (but fixed) stationary policy, and consider the transition kernel $Q_f$ in (2.3). As usual, the $n$-step transition kernel $(n = 1, 2, \ldots)$ is given by

$$Q_f^n(B \mid x) = \int_X Q_f^{n-1}(B \mid y)\, Q_f(dy \mid x),$$

where $Q_f^0(\cdot \mid x)$ is the Dirac measure $p_x$ concentrated at $x$.

Let $M_v$ be as in Definition 2.5. Then $Q_f$ defines a linear operator from $M_v$ to itself given by

$$Q_f\mu(\cdot) := \int_X Q_f(\cdot \mid x)\, \mu(dx), \qquad \mu \in M_v. \tag{3.1}$$

Moreover, using Assumption 2.4(b), a direct calculation (cf. Remark 2.10) shows that the operator norm of $Q_f$ satisfies

$$\|Q_f\|_v := \sup\{\|Q_f\mu\|_v : \|\mu\|_v \le 1\} \le \frac{\|\nu\|_v}{\overline{v}} + \alpha, \tag{3.2}$$

where $\overline{v}$ is the constant in Assumption 2.2(a).

LEMMA 3.2. *Suppose that Assumptions* 2.3 *and* 2.4 *hold. Then*, *when using a stationary policy $f \in \Delta_0$, the corresponding state process is aperiodic and uniformly ergodic with respect to the norm $\|\cdot\|_v$, the latter meaning that*

$$\|Q_f^t - \Pi_f\|_v \to 0 \qquad \text{as } t \to \infty, \tag{3.3}$$

*where $\Pi_f : M_v \to M_v$ denotes the stationary projector of $Q_f$, defined as*

$$\Pi_f\mu(\cdot) := \mu(X) q_f(\cdot) \qquad \forall \mu \in M_v. \tag{3.4}$$

P r o o f. By Kartashov's [12] Theorem E, it suffices to verify the following conditions (with $q_f, h_f$ and $\nu$ as in Assumptions 2.3 and 2.4):

(E.1) $\int_X h_f\, d\nu > 0$, $\int_X h_f\, dq_f > 0$;

(E.2) the kernel $\tau_f(B \mid x) := Q_f(B \mid x) - h_f(x)\nu(B)$ is nonnegative;

(E.3) $\|\tau_f\|_v := \sup\{\|\tau_f\mu\|_v : \|\mu\|_v \le 1\} \le \varrho$ for some $0 < \varrho < 1$.

The first inequality in (E.1) follows from Assumption 2.4(c), and the second from Proposition 5.6 in [18, p. 72]. Similarly, (E.2) follows from Assumption 2.4(a), and, finally, (E.3) holds with $\varrho = \alpha$ as in Assumption 2.4(b), since the latter yields

$$\begin{aligned}\|\tau_f\mu\|_v &= \int\int v(y)Q_f(dy \mid x)|\mu|(dx) - \Big(\int_X v\, d\nu\Big)\Big(\int_X h_f\, d|\mu|\Big)\\ &\le \|\nu\|_v \int_X h_f d|\mu| + \alpha\|\mu\|_v - \|\nu\|_v \int_X h_f d|\mu|\\ &\le \alpha \quad \text{if } \|\mu\|_v \le 1. \ \blacksquare\end{aligned}$$

Let $G$ be the function on $(0,1] \times (0,1]$ defined as

$$G(s,r) := \exp\Big(-\frac{1-s}{s}\cdot\frac{\log r}{1-r}\Big), \quad 0 < s, r \le 1. \tag{3.5}$$

Then from the conditions (E.1)–(E.3) in the proof of Lemma 3.1 and the Corollary to Theorem 6 in [13], we obtain the following estimate of the rate of convergence in (3.3):

Lemma 3.3. *Under Assumptions* 2.3 *and* 2.4, *for any* $f \in \Delta_0$,

$$\|Q_f^t - \Pi_f\|_v \le \Big(\alpha^t + \theta_0^{t+1}\frac{(t+2)e}{\alpha}\Big)(1+\sigma) \tag{3.6}$$

*for all* $t > \theta_0/(1-\theta_0)$, *where* $\sigma$ *and* $\theta_0$ *are positive constants satisfying*

$$\sigma \le \frac{b_0}{\overline{v}} \ \text{(with } b_0 \text{ as in Lemma 3.1(b))}, \tag{3.7}$$

$$\theta_0 = 1 - \frac{1-\alpha}{1+\alpha\sigma w} < 1, \quad \text{and} \tag{3.8}$$

$$w = 2G\Big(\int h_f\, dq_f, \int h_f\, d\nu\Big) - 1, \tag{3.9}$$

*with* $G$ *as in* (3.5) (*using Lemma* 3.1(a),(c)).

The next step is to use (3.6) to obtain an exponential rate of convergence of $Q_f^t$ to $\Pi_f$ *uniform* in $f \in \Delta_0$. Before stating Lemma 3.4, recall that $p_x$ denotes the Dirac measure concentrated at $x$, and note that

$$\|p_x\|_v = v(x) \quad \text{and} \quad Q_f^t p_x(\cdot) = Q_f^t(\cdot \mid x) = Q^t(\cdot \mid x, f). \tag{3.10}$$

Also observe that the proof of Lemma 3.4 shows how the constant $\overline{c}$ and $\eta$ appearing in the lemma can be estimated in terms of the quantities $\overline{v}$, $\alpha$, $\|\nu\|_v$ and $\gamma$ in Assumption 2.4.

LEMMA 3.4. *Under Assumptions* 2.3 *and* 2.4, *there exist positive constants* $\overline{c}$ *and* $\eta$, *with* $\eta < 1$, *such that for every* $x \in X$ *and* $t = 0, 1, \ldots,$

$$\sup_{\Delta_0} \|Q_f^t p_x - q_f\|_v \le \overline{c} v(x) \eta^t. \tag{3.11}$$

Proof. Let $G$ and $w$ as in (3.5), (3.9), and define

$$w_* := 2G(b_1, \gamma) - 1,$$

where $b_1$ and $\gamma$ are the constants in Lemma 3.1(a), (c). Thus $w \le w_*$ since $G(s, r)$ is decreasing in $s \in (0, 1]$ and $r \in (0, 1)$. Now let $b_0$ be as in Lemma 3.1(b) and define

$$\theta_* := 1 - \frac{1 - \alpha}{1 + \alpha w_* b_0 / \overline{v}} < 1.$$

Then (3.7)–(3.8), together with $w \le w_*$, imply $\theta_0 \le \theta_*$. Hence, since $w_*$ and $\theta_*$ are independent of $f \in \Delta_0$, the inequality (3.6) yields

$$\sup_{\Delta_0} \|Q_f^t - \Pi_f\|_v \le \left(\alpha^t + \frac{\theta_*^{t+1}(t+2)e}{\alpha}\right)\left(1 + \frac{b_0}{\overline{v}}\right) \tag{3.12}$$

for all $t > \theta_*/(1 - \theta_*)$. Now, let $t^* := [\theta_*/(1 - \theta_*) + 1]$, where $[r]$ stands for the integral part of $r$. Then, for any $t \le t_*$ and $f \in \Delta_0$, the inequality (3.2) and the fact that $\|\Pi_f\|_v \le b_0$ (see Lemma 3.1(b)) give

$$\|Q_f^t - \Pi_f\|_v \le \|Q_f\|_v^t + \|\Pi_f\|_v \le [\max(1, \|\nu\|_v/\overline{v} + \alpha)]^{t_*} + b_0, \tag{3.13}$$

where the right-hand side is independent of $f \in \Delta_0$. Therefore, from (3.12)–(3.13) we see that there are constants $\overline{c}$ and $\eta$, $\eta < 1$, satisfying

$$\sup_{\Delta_0} \|Q_f^t - \Pi_f\|_v \le \overline{c} \eta^t \quad \forall t = 0, 1, \ldots \tag{3.14}$$

Finally, to complete the proof of the lemma it suffices to note that

$$\|Q_f^t p_x - q_f\|_v = \|Q_f^t p_x - \Pi_f p_x\|_v \le \|Q_f^t - \Pi_f\|_v \|p_x\|_v,$$

which combined with (3.14) and (3.10) yields (3.11). ∎

Remark. Since $v(\cdot) \ge \overline{v}$ (see Assumption 2.2),

$$\|\mu\|_v \ge \overline{v} |\mu|(X) \quad \forall \mu \in M_v.$$

Thus, under Assumptions 2.3 and 2.4, Lemma 3.4 ensures geometric ergodicity of the transition kernels $Q_f$, $f \in \Delta_0$, in the weighted norm $\|\cdot\|_v$ and *also* in the total variation norm. Observe also (from (3.1)) that the geometric ergodicity is uniform in $f \in \Delta_0$, but is *not* uniform in the initial state $x \in X$. This situation is typical, e.g., in controlled queueing models.

To conclude this section we will use Lemma 3.4 to show that the expected cost $E_x^f c(x_t, f)$ (recall the notation (2.3b)) as a function of the initial state $x$ belongs to the space $\Phi_v$ (Definition 2.5) for all $f \in \Delta_0$ and $t = 0, 1 \ldots$ In precise terms we have:

LEMMA 3.5. *Under Assumptions* 2.2–2.4, *there is a constant* $c_1$ *such that*

$$\sup_t \sup_{\Delta_0} E_x^f c(x_t, f) \le c_1 v(x) \quad \forall x \in X. \tag{3.15}$$

Proof. For arbitrary $f \in \Delta_0$, $x \in X$ and $t \ge 0$,

$$E_x^f c(x_t, f) = \int c(y, f)\, Q^t(dy \mid x, f) \le \mathrm{I} + \mathrm{II},$$

where

$$\mathrm{I} := \Big| \int_X c(y, f)\, Q^t(dy \mid x, f) - \int_X c(y, f)\, q_f(dy) \Big|,$$

$$\mathrm{II} := \int_X c(y, f)\, q_f(dy).$$

Then, recalling the notation (3.10),

$$\begin{aligned} \mathrm{I} &\le \int_X c(y, f)\, |Q_f^t p_x - q_f|(dy) \le \int_X v(y)\, |Q_f^t p_x - q_f|(dy) \\ &= \|Q_f^t p_x - q_f\|_v \le \overline{c} v(x) \quad \text{(by Lemma 3.4, as } \eta < 1), \end{aligned} \tag{3.16}$$

and, similarly, by Lemma 3.1(b),

$$\mathrm{II} \le \int_X v(y) q_f(dy) = \|q_f\|_v \le b_0 \le \frac{b_0 v(x)}{\overline{v}}. \tag{3.17}$$

Thus from (3.16)–(3.17) we obtain (3.15) with $c_1 := \overline{c} + b_0/\overline{v}$. ∎

**4. Lemmas on discounted problems.** For every $\beta \in (0, 1)$, $x \in X$ and $\delta \in \Delta$, let

$$V_\beta(\delta, x) := E_x^\delta \Big[ \sum_{t=0}^{\infty} \beta^t c(x_t, a_t) \Big] = \lim_{n\to\infty} J_n^\beta(\delta, x) \tag{4.1}$$

be the total expected *β-discounted cost* (β-DC) when using the policy $\delta$, given the initial state $x_0 = x$, where (cf. (2.4))

$$J_n^\beta(\delta, x) := E_x^\delta \Big[ \sum_{t=0}^{n-1} \beta^t c(x_t, a_t) \Big]. \tag{4.2}$$

A policy $\delta^*$ is said to be *β-DC optimal* if

$$V_\beta(\delta^*, x) = \inf_\Delta V_\beta(\delta, x) =: V_\beta^*(x) \quad \forall x \in X; \tag{4.3}$$

$V_\beta^*$ is called the *optimal β-DC function*.

One of the main objectives in this section is to prove the following theorem, which is a well-known result under other sets of assumptions [2, 3, 6, 11].

THEOREM 4.1. *Suppose that Assumptions* 2.2–2.4 *hold*, *and let* $\beta \in (0,1)$ *be an arbitrary*, *but fixed*, *discount factor. Then*:

(a) $V_\beta^*$ *is the* (*pointwise*) *minimal solution in* $\Phi_v$ *of the* $\beta$*-discounted cost optimality equation* ($\beta$*-DCOE*)

$$V_\beta^*(x) = \min_{A(x)} \Big[ c(x,a) + \beta \int_X V_\beta^*(y)\, Q(dy \mid x,a) \Big], \qquad x \in X; \tag{4.4}$$

(b) *there exists a stationary policy* $f_\beta \in \Delta_0$ *such that* $f_\beta(x) \in A(x)$ *attains the minimum on the right-hand side of* (4.4) *for every* $x$ *in* $X$, *i.e.*,

$$V_\beta^*(x) = c(x, f_\beta) + \beta \int_X V_\beta^*(y)\, Q(dy \mid x, f_\beta), \tag{4.5}$$

*and* $f_\beta$ *is* $\beta$*-DC optimal.*

To prove Theorem 4.1 it is convenient to introduce some additional notation and preliminary results. *The hypotheses of Theorem* 4.1 *are supposed to hold throughout this section*.

Let $\Phi_v^+ := \{u \in \Phi_v : u \geq 0\}$ be the natural *positive cone* of $\Phi_v$, and for each positive number $\beta \leq 1$ and $u \in \Phi_v^+$ let $T_\beta u$ be the function on $X$ defined as

$$(T_\beta u)(x) := \inf_{A(x)} \Big[ c(x,a) + \beta \int_X u(y)\, Q(dy \mid x,a) \Big]. \tag{4.6}$$

The following lemma states that $T_\beta$ maps $\Phi_v^+$ into itself and that the infimum on the right-hand side of (4.6) is attained; hence we may write "minimum" instead of "infimum".

LEMMA 4.2. *For any* $u$ *in* $\Phi_v^+$ *and* $0 < \beta \leq 1$:

(a) *The function*

$$u'(x,a) := c(x,a) + \beta \int_X u(y)\, Q(dy \mid x,a), \qquad (x,a) \in \mathbb{K},$$

*is measurable on* $\mathbb{K}$ *and l.s.c. in* $a \in A(x)$ *for every* $x \in X$;

(b) *there exists* $f \in \Delta_0$ *such that*

$$\begin{aligned} T_\beta u(x) &= u'(x, f(x)) \\ &= c(x, f(x)) + \beta \int_X u(y)\, Q_f(dy \mid x) \qquad \forall x \in X, \end{aligned} \tag{4.7}$$

*and* $T_\beta u$ *is in* $\Phi_v^+$.

P r o o f. (a) First, by Assumption 2.2(a), if $u$ is in $\Phi_v^+$, then

$$0 \leq u'(x,a) \leq v(x) + \beta \|u\|_v \int_X v(y)\, Q(dy \mid x,a) < \infty, \tag{4.8}$$

so that $u'$ is a finite-valued function and its measurability follows from that of $c$ and the properties of the transition law $Q$. Now, to see that $a \mapsto u'(x,a)$ is l.s.c., let $u_n$ be a nondecreasing sequence of bounded functions such that $u_n \uparrow u$. Then if $\{a^l\}$ is a sequence in $A(x)$ converging to $a \in A(x)$, we have

$$\liminf_{l\to\infty} \int u(y)\,Q(dy \mid x,a^l) \ge \liminf_{l} \int u_n(y)\,Q(dy \mid x,a^l) = \int u_n(y)\,Q(dy \mid x,a) \quad \forall n,$$

by Assumption 2.2(c). Thus, letting $n\to\infty$, monotone convergence yields

$$\liminf_{l} \int u(y)\,Q(dy \mid x,a^l) \ge \int u(y)\,Q(dy \mid x,a),$$

i.e., $a \mapsto \int u(y)\,Q(dy \mid x,a)$ is l.s.c. on $A(x)$ for every $x \in X$, which combined with the l.s.c. of $a \mapsto c(x,a)$ completes the proof of part (a).

(b) The existence of a "minimizer" $f \in \Delta_0$ satisfying (4.7) follows from (a) and Corollary 4.3 (and the remark following it) in [21]. The fact that $T_\beta u$ is in $\Phi_v^+$ follows from (4.7)–(4.8) and Assumption 2.4(b). ■

For each $0<\beta\le 1$, let $\{v_n^\beta\}$ be the sequence of *value iteration* functions defined recursively as

(4.9) $$v_n^\beta(x) := T_\beta v_{n-1}^\beta(x), \qquad x\in X,\ n=1,2,\ldots,$$

with $v_0^\beta(\cdot)\equiv 0$. An elementary induction argument and Lemma 4.2 yield the following.

LEMMA 4.3. *For every $0<\beta\le 1$ and $n=1,2,\ldots$, $v_n^\beta$ is in $\Phi_v^+$ and there exists $f_n^\beta\in\Delta_0$ such that, for all $x\in X$,*

(4.10) $$v_n^\beta(x) = \min_{A(x)}\Big[c(x,a)+\beta\int_X v_{n-1}^\beta(y)\,Q(dy\mid x,a)\Big] = c(x,f_n^\beta)+\beta\int_X v_{n-1}^\beta(y)\,Q(dy\mid x,f_n^\beta). \ ■$$

Moreover, from elementary stochastic dynamic programming [2, 3, 9, 11], $v_n^\beta$ is the *optimal n-stage cost*, i.e. (see (4.2)),

(4.11) $$v_n^\beta(x) = \inf_{\Delta} J_n^\beta(\delta,x).$$

This implies that for any positive $\beta<1$,

$$v_n^\beta(x)\le J_n^\beta(\delta,x)\le V_\beta(\delta,x) \quad \forall\delta\in\Delta,\ x\in X,$$

which in turn yields

$$v_n^\beta(x)\le V_\beta^*(x) \quad \forall x\in X.$$

Thus, since $T_\beta$ is monotone (i.e., $u\ge u' \Rightarrow T_\beta u\ge T_\beta u'$), the functions $v_n^\beta$ form a nondecreasing sequence converging to a function $u\le V_\beta^*$. Therefore, to complete the *proof of Theorem* 4.1, one can use standard arguments (see e.g. [2, 8, 9]) to show that: (i) $u$ satisfies the $\beta$-DCOE (4.4), (ii) $u$ is the minimal, measurable, nonnegative solution of (4.4), and (iii) $u = V_\beta^*$.

On the other hand, by Lemma 3.5, $V_\beta^*$ is indeed in $\Phi_v$, as (3.15) gives

$$V_\beta^*(x) \le \frac{c_1 v(x)}{1-\beta} \quad \forall x \in X. \tag{4.12}$$

This completes the proof of Theorem 4.1(a), and, finally, part (b) follows from Lemma 4.2 (with $0 < \beta < 1$). ∎

We next introduce two auxiliary functions to be used in the proof of Theorem 2.6.

DEFINITION 4.4. Let $z \in X$ be an arbitrary, fixed state, and for every $0 < \beta < 1$ and $x \in X$, let

$$\phi_\beta(x) := V_\beta^*(x) - V_\beta^*(z), \qquad j_\beta := (1-\beta) V_\beta{}^*(z). \tag{4.13}$$

Observe that the $\beta$-DCOE (4.4) can be written, equivalently, as

$$j_\beta + \phi_\beta(x) = \min_{A(x)} \Big[ c(x,a) + \beta \int_X \phi_\beta(y)\, Q(dy \mid x, a) \Big]. \tag{4.14}$$

Similarly, we may rewrite (4.5) as

$$j_\beta + \phi_\beta(x) = c(x, f_\beta) + \beta \int_X \phi_\beta(y)\, Q(dy \mid x, f_\beta). \tag{4.15}$$

We also have:

LEMMA 4.5. *There are constants* $c_1$ *and* $c_2$ *such that*, *for every* $\beta$ *in* $(0,1)$ *and* $x \in X$,

(a) $0 \le j_\beta \le c_1 v(z)$, *and*
(b) $|\phi_\beta(x)| \le c_2(1 + v(z)/\overline{v}) v(x)$, *i.e.*, $\phi_\beta$ *is in* $\Phi_v$.

Proof. (a) follows from (4.12).

(b) As a consequence of Theorem 4.1, to find a $\beta$-DC optimal policy we may restrict ourselves to the class $\Delta_0$ of stationary policies, i.e., we may write (4.3) as

$$V_\beta^*(x) = \inf_{\Delta_0} V_\beta(f, x).$$

Thus

$$\begin{aligned} |\phi_\beta(x)| &= |\inf_{\Delta_0} V_\beta(f,x) - \inf_{\Delta_0} V_\beta(f,z)| \le \sup_{\Delta_0} |V_\beta(f,x) - V_\beta(f,z)| \\ &\le \sup_{\Delta_0} \sum_{t=0}^{\infty} \beta^t |E_x^f c(x_t, f) - E_z^f c(x_t, f)|. \end{aligned} \tag{4.16}$$

To estimate the right-hand side of (4.16), inside the absolute value add and subtract $\int_X c(y,f)\,q_f(dy)$, and then use (2.2a) and (3.11) to obtain

$$|E_x^f c(x_t,f) - E_z^f c(x_t,f)| \le \|Q_f^t p_x - q_f\|_v + \|Q_f^t p_z - q_f\|_v \le \overline{c}\eta^t[v(x)+v(z)].$$

Thus, since $v(x)+v(z) \le (1+v(z)/\overline{v})v(x)$ and $\beta < 1$, (4.16) yields

$$|\phi_\beta(x)| \le \overline{c}(1-\eta)^{-1}\left(1+\frac{v(z)}{\overline{v}}\right)v(x),$$

and (b) follows with $c_2 := \overline{c}(1-\eta)^{-1}$, which is a constant independent of $\beta$ and $x$. ■

**5. Proof of Theorem 2.6.** After the preliminary results in Sections 3 and 4, the proof of Theorem 2.6 is similar, *mutatis mutandis*, to the proof of Theorem 4.2 in [8]. To begin with, Lemma 4.5(a) ensures the existence of a number $\varrho^* \ge 0$ such that

$$\limsup_{\beta\uparrow 1} j_\beta = \varrho^*,$$

where $j_\beta := (1-\beta)V_\beta^*(z)$; see (4.13). Let $\{\beta_n\}$ be a sequence of discount factors such that $\beta_n \uparrow 1$ and

$$\lim_{n\to\infty}(1-\beta_n)V_{\beta_n}^*(z) = \varrho^*. \tag{5.1}$$

The following lemma shows that (5.1) holds if $z$ is replaced by any state $x \in X$.

Lemma 5.1. $\lim_{n\to\infty}(1-\beta_n)V_{\beta_n}^*(x) = \varrho^*$ *for all* $x$ *in* $X$.

Proof. Since

$$|(1-\beta_n)V_{\beta_n}^*(x) - \varrho^*| \le (1-\beta_n)|\phi_{\beta_n}(x)| + |(1-\beta_n)V_{\beta_n}^*(z) - \varrho^*|,$$

the desired result follows from (5.1) and Lemma 4.5(b). ■

Moreover, by a well-known Tauberian theorem (cf. [25] or [1, 8, 9]) the optimal AC-function $J^*(x)$ satisfies

$$\limsup_{\beta\uparrow 1}(1-\beta)V_\beta^*(x) \le J^*(x) \quad \forall x \in X;$$

hence, by Lemma 5.1,

$$\varrho^* \le J^*(x) \quad \forall x \in X. \tag{5.2}$$

We also have:

Lemma 5.2. *Suppose that there exist a constant* $\varrho^* \ge 0$, *a function* $\phi^* \in \Phi_v$ *and a stationary policy* $f^* \in \Delta_0$ *such that*

$$\varrho^* + \phi^*(x) \ge c(x,f^*) + \int_X \phi^*(y)\,Q(dy \mid x, f^*) \quad \forall x \in X. \tag{5.3}$$

*Then* $f^*$ *and* $\varrho^*$ *satisfy Theorem* 2.6(b), *i.e.*,

$$J(f^*, x) = J^*(x) = \varrho^*. \tag{5.4}$$

Proof. Iteration of (5.3) yields, for all $n \ge 1$ and $x \in X$,

$$n\varrho^* + \phi^*(x) \ge J_n(f^*, x) + \int_X \phi^*(x)\, Q^n(dy \mid x, f^*). \tag{5.5}$$

On the other hand, by Lemma 3.4,

$$\begin{aligned}(5.6)\quad & \Big| \int \phi^*(y)\, Q^n(dy \mid x, f^*) - \int \phi^*(y)\, q_{f^*}(dy) \Big| \\ & \qquad \le \|\phi^*\|_v \|Q^n_{f^*} p_x - q_{f^*}\|_v \le \|\phi^*\|_v \overline{c} v(x) \eta^n \to 0 \quad \text{as } n \to \infty.\end{aligned}$$

Thus, dividing by $n$ in (5.5) and letting $n \to \infty$, we obtain

$$\varrho^* \ge J(f^*, x) \quad \forall x \in X.$$

Finally, since (by definition of $J^*$ and (5.2))

$$\varrho^* \le J^*(x) \le J(f^*, x) \quad \forall x \in X,$$

we obtain (5.4). ■

It follows from Lemma 5.2 that the *proof of Theorem* 2.6 will be complete if we can show the existence of a triplet $(\varrho^*, \phi^*, f^*)$, with $\varrho^* \ge 0$, $\phi^*$ in $\Phi_v$, and $f^*$ in $\Delta_0$, satisfying (2.8). To obtain this, let $\varrho^*$ be as in Lemma 5.1, and define

$$\phi^*(x) := \liminf_{n \to \infty} \phi_{\beta_n}(x), \quad x \in X. \tag{5.7}$$

By Lemma 4.5(b), the function $\phi^*$ is in $\Phi_v$. Finally, let $f_\beta \in \Delta_0$ be the $\beta$-DC $(0 < \beta < 1)$ optimal policy in (4.14)–(4.15) (see also Theorem 4.1) and consider the sequence $\{f_{\beta_n}\}$. By Schäl's [23, Proposition 12.2], there is a stationary policy $f^*$ such that $f^*(x) \in A(x)$ is an accumulation point of $\{f_{\beta_n}(x)\}$ for each $x$ in $X$. Then a straightforward modification in the proof of [8, Theorem 4.2]—replacing $\phi_{\beta_n}(x)$ by the nonnegative function $\phi_{\beta_n}(x) + c_3 v(x) \ge 0$, where $c_3 := c_2(1 + v(z)/\overline{v})$ (see Lemma 4.5(b)), and using (2.2c)—shows that $(\varrho^*, \phi^*, f^*)$ thus defined satisfies the ACOI (2.8). ■

Remark 5.3. Note that, from (5.7) and (4.13), $\phi^*(z) = 0$.

**6. Proof of Theorem 2.8.** Throughout this section we suppose that the hypotheses of Theorem 2.8 (i.e., Assumptions 2.2, 2.3, 2.4 and 2.7) hold true.

Let $\phi_\beta$ be as in (4.13). The first step in the proof of Theorem 2.8 is to show that:

(6.1) *the family of functions* $\{\phi_\beta : 0 < \beta < 1\}$ *is equicontinuous.*

This will follow from (4.14) and Lemma 4.5(b), together with the following general result in which $\Psi$ is the class of functions introduced in the paragraph preceding Assumption 2.7. We also use the following *notation*: If $u \in \Phi_v$, $0 < \beta \le 1$ and $b > 0$:

$$U_\beta(x) := \inf_{A(x)} \Big[ c(x,a) + \beta \int_X u(y)\, Q(dy \mid x,a) \Big], \qquad x \in X,$$
$$G_b := \{u \in \Phi_v : \|u\|_v \le b\}.$$

LEMMA 6.1. *If Assumption* 2.7 *holds*, *then for each given* $b > 0$ *there exists a family of functions* $\{\psi_x : x \in X\}$ *in* $\Psi$ *such that for every* $u \in G_b$, $x \in X$ *and* $0 < \beta \le 1$,

$$|U_\beta(x) - U_\beta(x')| \le \psi_x[d_1(x,x')] \qquad \forall x' \in X, \tag{6.2}$$

*where* $d_1$ *is the metric on* $X$.

Proof. Let $u$ be an arbitrary function in $G_b$ and define

$$u'(x,a) := c(x,a) + \beta \int_X u(y)\, Q(dy \mid x,a), \qquad (x,a) \in \mathbb{K},$$

so that $U_\beta(x) = \inf_{A(x)} u'(x,a)$. Now, by Assumption 2.7(b), and writing $k := (x,a)$, $k' := (x',a')$,

$$\begin{aligned}|u'(k) - u'(k')| &\le |c(k) - c(k')| + \beta \int |u(y)||Q(dy \mid k) - Q(dy \mid k')| \\ &\le \psi_x^c[d(k,k')] + \beta\|u\|_v \psi_x^Q[d(k,k')] \le \widehat{\psi}_x[d(k,k')],\end{aligned}$$

where $\widehat{\psi}_x := \psi_x^c + b\psi_x^Q$ is a function in $\Psi$. This yields—as in the proof of Gordienko's [4, Lemma 1]—the existence of functions $\psi_x$ in $\Psi$ satisfying (6.2), which proves the lemma. ∎

As already noted, (4.14), Lemma 4.5(b) and Lemma 6.1 together imply (6.1). Now let $\{\beta_n\}$ be the sequence in (5.1). Then, by (6.1) and the Ascoli Theorem (see e.g. Royden [22, p. 179]), there exists a subsequence of $\{\phi_{\beta_n}\}$ (also denoted by $\{\phi_{\beta_n}\}$) and a continuous function $\phi^*$ on $X$ such that

$$\lim_{n\to\infty} \phi_{\beta_n}(x) = \phi^*(x) \qquad \forall x \in X, \tag{6.3}$$

the convergence being uniform on compact subsets of $X$. Moreover, by Lemma 4.5(b), $\phi^*$ is in $\Phi_v$, and, on the other hand, by (6.3), $\{\phi_{\beta_n}\}$ satisfies (5.7). Therefore, all the conclusions of Theorem 2.6 hold, of course, in the present case.

Thus, to complete the proof of Theorem 2.8 it only remains to show that

$$\varrho^* + \phi^*(x) \le \min_{A(x)} \Big[ c(x,a) + \int_X \phi^*(y)\, Q(dy \mid x,a) \Big]; \tag{6.4}$$

cf. (2.8) and (2.10). To prove (6.4) note that, from (4.14),

$$j_{\beta_n} + \phi_{\beta_n}(x) \le c(x,a) + \beta_n \int_X \phi_{\beta_n}(y)\, Q(dy \mid x,a)$$

for every $x \in X$ and $a \in A(x)$. Finally, letting $n \to \infty$, (6.3), (2.2b) and the Dominated Convergence Theorem yield

$$\varrho^* + \phi^*(x) \le c(x,a) + \int_X \phi^*(y)\, Q(dy \mid x,a),$$

which implies (6.4). ∎

**7. Corollaries and concluding remarks.** In this section we state two important consequences of Theorems 2.6 and 2.8. The first one was in fact proved in the last paragraph of Section 5.

COROLLARY 7.1. *Under the assumptions of Theorem* 2.6, *there exists a sequence of discount factors* $\beta_n \uparrow 1$, *a sequence* $\{f_{\beta_n}\}$ *of* $\beta_n$*-DC optimal stationary policies, and an AC-optimal stationary policy* $f^*$ *such that for every state* $x$, $f^*(x)$ *is an accumulation point of* $\{f_{\beta_n}(x)\}$.

In short, Corollary 7.1 states that there is an AC-optimal policy which is an "accumulation point" of DC-optimal policies as the discount factor increases to 1.

To state the second corollary we need some additional notation: Let $J_n(\delta,x)$ be as in (2.4) and, given a measurable function $h : X \to \mathbb{R}$, let

$$J_n(\delta,x,h) := E_x^\delta\Big[\sum_{t=0}^{n-1} c(x_t,a_t) + h(x_n)\Big] = J_n(\delta,x) + E_x^\delta h(x_n) \tag{7.1}$$

be the expected $n$-stage cost when using the policy $\delta$, given the initial state $x$, and the *terminal cost* function $h$. Let

$$J_n^*(x,h) := \inf_{\Delta} J_n(\delta,x,h) \tag{7.2}$$

be the corresponding $n$-stage *optimal cost*, and if $h(\cdot) \equiv 0$ write

$$J_n(\delta,x,0) := J_n(\delta,x), \qquad J_n^*(x,0) = J_n^*(x). \tag{7.3}$$

Then, by a well-known characterization of canonical triplets—see e.g. [16, Theorem 3.2] (for *bounded* one-stage costs $c$, see [3, pp. 166–167], or [1, Theorem 6.2])—we conclude that Theorem 2.8 can be restated as follows.

COROLLARY 7.2. *Under the assumptions of Theorem* 2.8, *there is a canonical triplet* $(\varrho^*,\phi^*,f^*)$ *such that*

$$J_n(f^*,x,\phi^*) = J_n^*(x,\phi^*) = n\varrho^* + \phi^*(x) \tag{7.4}$$

*for every* $x \in X$ *and* $n = 1,2,\ldots$

In other words, the first equality in (7.4) states that, for every $n \geq 1$ and every initial state $x_0 = x$, $f^* \in \Delta_0$ is an *optimal policy* for the $n$-stage problem with terminal cost function $\phi^*$, whereas the second equality says that the corresponding *optimal cost* is $n\varrho^* + \phi^*(x)$.

In a companion paper [5] we show that the policy $f^*$ in Theorem 2.8 and Corollary 7.2 is in fact "AC-optimal" in several ways *stronger* than the AC-optimality defined in Section 2, and also that the optimal (constant) AC-function $\varrho^*$ may be obtained by the so-called *value iteration* (or *successive approximations*) procedure. The latter yields, in particular, that

$$\lim_{n\to\infty} \frac{J_n^*(x)}{n} = \varrho^* \qquad \forall x \in X,$$

where $J_n^*(x)$ is the function in (7.3), so that $\varrho^*$ can be obtained as the limit of "averaged" finite ($n$-stage) horizon problems. Moreover, among additional results, [5] presents an example in which all the hypotheses of Theorem 2.8 are satisfied.

**8. Appendix.** Let $X$ be the Borel (state) space of Section 2, and let $\{x_t : t = 0, 1, \ldots\}$ be an $X$-valued (non-controlled) Markov process with transition kernel $P(B \mid x)$, with $B \in \mathbb{B}_X$, $x \in X$. This process—or the transition kernel—is said to be *Harris-recurrent* if there exists a nontrivial $\sigma$-finite measure $\lambda$ on $X$ such that

$$P(x_t \in B \text{ for some } t \mid x_0 = x) = 1 \qquad \forall x \in X$$

whenever $B \in \mathbb{B}_X$ satisfies $\lambda(B) > 0$ [10, 20]. The following proposition is well known [15, 18, 19].

PROPOSITION 8.1. *If* $\{x_t\}$ *is Harris-recurrent, then*:

(a) *there exists a nontrivial*, $\sigma$*-finite invariant measure* $\pi$ *for the transition kernel* $P$; *moreover*, $\pi$ *is unique up to a multiplicative constant*;

(b) *there exists a triplet* $(n, \nu, h)$ *consisting of an integer* $n \geq 1$, *a probability measure* $\nu$, *and a nonnegative function* $h \leq 1$ *such that*

(i) $P^n(B \mid x) \geq h(x)\nu(B)$ *for* $B \in \mathbb{B}_X$ *and* $x \in X$,
(ii) $\int_X h\, d\nu > 0$, *and*
(iii) $0 < \int_X h\, d\pi < \infty$.

The Markov process is called *positive Harris-recurrent* if it is Harris-recurrent and it has an invariant probability measure (cf. Proposition 8.1(a)).

## References

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

[2] D. P. Bersekas and S. E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.

[3] E. B. Dynkin and A. A. Yushkevich, *Controlled Markov Processes*, Springer, New York, 1979.

[4] E. I. Gordienko, *Controlled Markov processes with slowly varying characteristics. The problem of adaptive control. I*, Soviet J. Comput. Syst. Sci. 23 (1985), 87–95.

[5] E. I. Gordienko and O. Hernández-Lerma, *Average cost Markov control processes with weighted norms: value iteration*, this volume, 219–237.

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

[7] —, *Average optimality in dynamic programming on Borel spaces—unbounded costs and controls*, Systems Control Lett. 17 (1991), 237–242.

[8] O. Hernández-Lerma and J. B. Lasserre, *Average cost optimal policies for Markov control processes with Borel state space and unbounded costs*, ibid. 15 (1990), 349–356.

[9] —, —, *Discrete-Time Markov Control Processes*, book in preparation.

[10] O. Hernández-Lerma, R. Montes-de-Oca and R. Cavazos-Cadena, *Recurrence conditions for Markov decision processes with Borel state space: a survey*, Ann. Oper. Res. 28 (1991), 29–46.

[11] K. Hinderer, *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*, Lecture Notes Oper. Res. 33, Springer, New York, 1970.

[12] N. V. Kartashov, *Inequalities in theorems of ergodicity and stability of Markov chains with common phase space. I*, Theory Probab. Appl. 30 (1985), 247–259.

[13] —, *Inequalities in theorems of ergodicity and stability of Markov chains with common phase space. II*, ibid. 30 (1985), 507–515.

[14] —, *Strongly stable Markov chains*, J. Soviet Math. 34 (1986), 1493–1498.

[15] V. K. Malinovskiĭ, *Limit theorems for Harris Markov chains, I*, Theory Probab. Appl. 31 (1986), 269–285.

[16] R. Montes-de-Oca and O. Hernández-Lerma, *Conditions for average optimality in Markov control processes with unbounded costs and controls*, J. Math. Systems Estim. Control 4 (1994), 1–19.

[17] —, —, *Value iteration in average cost Markov control processes on Borel spaces*, Acta Appl. Math., to appear.

[18] E. Nummelin, *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge University Press, Cambridge, 1984.

[19] E. Nummelin and P. Tuominen, *Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory*, Stochastic Process. Appl. 12 (1982), 187–202.

[20] S. Orey, *Limit Theorems for Markov Chain Transition Probabilities*, Van Nostrand Reinhold, London, 1971.

[21] U. Rieder, *Measurable selection theorems for optimization problems*, Manuscripta Math. 24 (1978), 115–131.

[22] H. L. Royden, *Real Analysis*, 2nd ed., Macmillan, New York, 1971.

[23] M. Schäl, *Conditions for optimality and for the limit of n-stage optimal policies to be optimal*, Z. Wahrsch. Verw. Gebiete 32 (1975), 179–196.

[24] —, *Average optimality in dynamic programming with general state space*, Math. Oper. Res. 18 (1993), 163–172.

[25] R. Sznajder and J. A. Filar, *Some comments on a theorem of Hardy and Littlewood*, J. Optim. Theory Appl. 75 (1992), 201-209.

EVGUENI GORDIENKO
DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD AUTÓNOMA METROPOLITANA-I
APARTADO POSTAL 55-534
09340 MÉXICO D.F., MEXICO

ONÉSIMO HERNÁNDEZ-LERMA
DEPARTAMENTO DE MATEMÁTICAS
CINVESTAV-IPN
APARTADO POSTAL 14-740
07000 MÉXICO D.F., MEXICO
E-mail: OHERNAND@MATH.CINVESTAV.MX

*Received on 2.10.1994*