J. Appl. Prob. **43,** 318–334 (2006)
Printed in Israel

AVERAGE OPTIMALITY FOR MARKOV DECISION PROCESSES IN BOREL SPACES: A NEW CONDITION AND APPROACH

XIANPING GUO,* *Zhongshan University*
QUANXIN ZHU,** *South China Normal University*

Abstract

In this paper we study discrete-time Markov decision processes with Borel state and action spaces. The criterion is to minimize average expected costs, and the costs may have *neither upper nor lower* bounds. We first provide *two* average optimality inequalities of opposing directions and give conditions for the existence of solutions to them. Then, using the two inequalities, we ensure the existence of an average optimal (deterministic) stationary policy under additional continuity–compactness assumptions. Our conditions are slightly *weaker* than those in the previous literature. Also, some *new* sufficient conditions for the existence of an average optimal stationary policy are imposed on the *primitive data* of the model. Moreover, our approach is slightly different from the well-known 'optimality inequality approach' widely used in Markov decision processes. Finally, we illustrate our results in two examples.

Keywords: Discrete-time Markov decision process; average expected criterion; average optimality inequality; optimal stationary policy

2000 Mathematics Subject Classification: Primary 90C40; 93E20

1. Introduction

The long-run *average expected criterion* in discrete-time Markov decision processes has been widely studied. As is well known, when the state and action spaces are both *finite*, the existence of an average optimal stationary policy is indeed guaranteed [7, pp. 165–176], [6, p. 71], [18, p. 450]. However, when a state space is countably infinite, an average optimal policy may *not* exist even though the action space is compact [7, p. 178], [18, p. 413]. Thus, the main goal has been to find optimality conditions (i.e. conditions for the existence of an average optimal policy). Much work on this has been done; for instance, see [1], [2], [5], [11], [18, pp. 414–416], [25], and [24, 132–135] for denumerable Markov decision processes and [2], [7, p. 188], [8], [9], [10], [14, pp 67–69], [13, p. 86], [12, p. 128], [15], and [16] for Markov decision processes in Borel spaces. In this paper, we will deal with the average expected criterion for Markov decision processes in Borel spaces, so here we describe some existing works on Markov decision processes in Borel spaces.

Received 9 September 2004; revision received 13 March 2006.
* Postal address: School of Mathematics and Computational Science, Zhongshan University, Guangzhou, 510275, PR China. Email address: mcsgxp@mail.sysu.edu.cn
** Postal address: Department of Mathematics, South China Normal University, Guangzhou, 510631, PR China. Email address: zqx1975@sina.com.cn
Partially supported by the NSFC, the NCET, and the RFDP.

(i) For costs/rewards that are *bounded*, the *minorant condition* for the existence of both a bounded solution to the average optimality equation (AOE) and an average optimal stationary policy was given in [7, p. 188]. The main results of [7] have been extended to the case with *ergodicity conditions* in [9], [10], and [12, p. 56]. The methods used there to ensure the existence of a bounded solution to the AOE employ *Banach's fixed-point theorem*.

(ii) When the costs are *nonnegative* (or bounded below), one of the optimality conditions is that the *relative difference*, $h_\alpha(x) := V_\alpha^*(x) - V_\alpha^*(x_0)$, of the discounted optimal value function, $V_\alpha^*(x)$, is assumed to be bounded below in both state, x, and discount factor, α, and the *optimality inequality approach* used to prove the existence of an average optimal stationary policy employs the Abelian theorem relating the average cost criterion to the discounted cost criterion; see [2], [14, p. 128], and [13, p. 86], for instance. It should be noted that, in order to use the Abelian theorem, the costs have to be nonnegative (or bounded below). Thus, the optimality inequality approach above is *not* applicable when the costs have neither upper nor lower bounds.

(iii) For the much more general case in which the costs have *neither upper nor lower* bounds, in order to establish the AOE and then prove the existence of an average optimal stationary policy, the equicontinuity condition of $h_\alpha(x)$ [8], [13, p. 96] or the irreducibility condition (e.g. Assumption 10.3.5 of [14, p. 130]) is required. Also, under the slightly stronger condition that transition laws have transition densities satisfying continuity–compactness, uniform ergodicity, and uniform integrability hypotheses, stronger results (e.g. a Blackwell optimal policy) have been established, in [15] and [16].

In this paper we study the general case further. We not only give *another set* of optimality conditions slightly weaker than those in the previous literature, but also provide a *new* approach to prove the existence of an average optimal stationary policy. More precisely, we require that the function $h_\alpha(x)$ is bounded *only* in the discount factor and remove both the equicontinuity condition of $h_\alpha(x)$ used in [8] and [13, p. 96] and the irreducibility condition used in [14, p. 130]. Thus, we can neither ensure the existence of a solution to the AOE nor use the Abelian theorem since in our model the state space may not be denumerable, the irreducibility condition of [14] has been removed, and the costs may have neither upper nor lower bounds. Instead, we first use *two* average optimality inequalities to replace the AOE used in [8], [14], and [13] and ensure that solutions to them exist. Then we prove the existence of an average optimal stationary policy using the two inequalities. Moreover, following the ergodicity conditions in [1], [5], [14, p. 122], [15], and [16], we give *new* sufficient conditions for our assumptions to hold. These conditions are imposed on the controlled system's *primitive data*.

Finally, we use a generalized inventory system to show that all conditions in this paper are satisfied, whereas some of the conditions in [1], [2], [3], [4], [7, p. 184], [5], [14, p. 128], [13, p. 46], [12, p. 7], [18, p. 18], [19], [22], [20], [25], and [24, p. 15] fail to hold, and we also apply our results to a controlled queueing model. It should be mentioned that the conditions of [15] and [16] also apply to the generalized inventory system. It is a very interesting, and so far unsolved, problem to find a real model for which all conditions in this paper are fulfilled, but which does not satisfy the assumptions made in [15] and [16].

The rest of the paper is organized as follows. In Section 2 we introduce our control model and the optimality problem. After giving optimality conditions and some technical preliminaries in Section 3, we study the existence of an average optimal stationary policy in Section 4. In Section 5 we use two examples to illustrate our results. We conclude in Section 6 with some general remarks.

2. The optimal control problem

In this section we first introduce the control model,

$$\{S, (A(x) \subset A,\ x \in S), Q(\cdot \mid x, a), c(x, a)\},$$

where S and A are respectively the state and action spaces, which are assumed to be Borel spaces, and $A(x)$ denotes the set of available actions at state $x \in S$. Suppose that the set

$$K := \{(x, a) \colon x \in S,\ a \in A(x)\}$$

is also a Borel space. The function $Q(\cdot \mid x, a)$ with $(x, a) \in K$, the *transition law*, is a stochastic kernel on S given K. Finally, $c(x, a)$ with $(x, a) \in K$, the *cost function*, is assumed to be a real-valued and measurable function on K. (As $c(x, a)$ is allowed to take both positive and negative values, it can also be interpreted as a *reward function*.)

To introduce the optimal control problem that we are concerned with, we need to introduce the policy classes. For each $t \geq 0$, let H_t be the family of admissible histories up to time t, that is, $H_0 := S$ and $H_t := K \times H_{t-1}$ for each $t \geq 1$.

Definition 2.1. A *randomized history-dependent policy* is a sequence $\pi := (\pi_t,\ t \geq 0)$ of stochastic kernels π_t on A given H_t that satisfy

$$\pi_t(A(x) \mid h_t) = 1 \quad \text{for all } h_t = (x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x) \in H_t \text{ and } t \geq 0.$$

The class of all randomized history-dependent policies is denoted by Π. A randomized history-dependent policy $\pi := (\pi_t,\ t \geq 0) \in \Pi$ is called *stationary* if there exists a measurable function f on S, with $f(x) \in A(x)$ for all $x \in S$, such that

$$\pi_t(\{f(x)\} \mid h_t) = \pi_t(\{f(x)\} \mid x) = 1$$
$$\text{for all } h_t = (x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x) \in H_t \text{ and } t \geq 0.$$

For simplicity, we denote this stationary policy by f. The class of all stationary policies is denoted by F, which means that F is the set of all measurable functions f on S with $f(x) \in A(x)$ for all $x \in S$.

If X is a Borel space, we denote by $\mathcal{B}(X)$ its Borel σ-algebra.

For each $x \in S$ and $\pi \in \Pi$, by the well-known Tulcea theorem (see [7, p. 16], [14, p. 42], and [12, p. 4]), there exist a unique probability measure space $(\Omega, \mathcal{F}, \mathrm{P}_x^\pi)$ and discrete-time stochastic processes $\{x_t\}$ and $\{a_t\}$, defined on Ω, such that, for each $D \in \mathcal{B}(S)$ and $t \geq 1$,

$$\mathrm{P}_x^\pi(x_{t+1} \in D \mid h_t, a_t) = Q(D \mid x_t, a_t) \tag{2.1}$$

for $h_t = (x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t) \in H_t$, where x_t and a_t denote the state and action variables at time $t \geq 1$, respectively. The expectation operator with respect to P_x^π is denoted by E_x^π. In particular, when the policy $\pi := f$ is in F, the *corresponding process*, $\{x_t\}$ (with values in S), is a Markov chain with transition law $Q(\cdot \mid x, f(x))$.

Now we define the long-run average cost criterion, $\bar{V}(\cdot, \cdot)$, and the corresponding optimal value function, $\bar{V}^*(\cdot)$. For each $x \in S$ and $\pi \in \Pi$,

$$\bar{V}(x, \pi) := \limsup_{n \to \infty} \frac{\mathrm{E}_x^\pi[\sum_{t=0}^{n-1} c(x_t, a_t)]}{n}, \qquad \bar{V}^*(x) := \inf_{\pi \in \Pi} \bar{V}(x, \pi).$$

A policy $\pi^* \in \Pi$ is called average optimal if $\bar{V}(x, \pi^*) = \bar{V}^*(x)$ for all $x \in S$.

The main aim of this paper is to give new conditions for the existence of an average optimal stationary policy.

https://doi.org/10.1239/jap/1152413725 Published online by Cambridge University Press

3. Optimality conditions

In this section we state conditions for the existence of an average optimal stationary policy and give some preliminary lemmas that are needed to prove our main results.

Since the cost function, $c(x, a)$, may be unbounded, to guarantee the finiteness of $\bar{V}(x, \pi)$ we first use the 'expected growth' condition, (3.1), below.

Assumption 3.1. (i) *There exist positive constants, b and $\beta < 1$, and a (measurable) function, $w \geq 1$, on S such that*

$$\int_S w(y)Q(\mathrm{d}y \mid x, a) \leq \beta w(x) + b \quad \textit{for all } (x, a) \in K. \tag{3.1}$$

(ii) *There exists a constant, $M > 0$, such that $|c(x, a)| \leq Mw(x)$ for all $(x, a) \in K$.*

Remark 3.1. Assumption 3.1(i) is well known as the statement of the 'Lyapunov-like inequality'; see [14, p. 121], for instance. Obviously, the constant b in (3.1) can be replaced by a *bounded* nonnegative measurable function, $b(x)$, on S as in Assumption 10.2.1(f) of [14, p. 121].

Lemma 3.1. *Suppose that Assumption 3.1 holds. Then*

(a) $\mathrm{E}_x^\pi[w(x_t)] \leq \beta^t w(x) + [(1 - \beta^t)/(1 - \beta)]b$ *for all $t \geq 0$, $x \in S$, and $\pi \in \Pi$; and*

(b) $|\bar{V}(x, \pi)| \leq bM/(1 - \beta)$ *for all $x \in S$ and $\pi \in \Pi$.*

Proof. (a) We prove Lemma 3.1(a) by induction. It is obviously valid for $t = 0$. For any $t \geq 1$, by Assumption 3.1(i) and (2.1) we have

$$\mathrm{E}_x^\pi[w(x_t) \mid x_0, a_0, x_1, a_1, \ldots, x_{t-1}, a_{t-1}] = \int_S w(y)Q(\mathrm{d}y \mid x_{t-1}, a_{t-1}) \leq \beta w(x_{t-1}) + b$$

and, so,

$$\begin{aligned}\mathrm{E}_x^\pi[w(x_t)] &\leq \beta\, \mathrm{E}_x^\pi[w(x_{t-1})] + b \\ &\leq \beta^2\, \mathrm{E}_x^\pi[w(x_{t-2})] + b + b\beta \leq \cdots \leq \beta^t w(x) + b + b\beta + \cdots + b\beta^{t-1} \\ &= \beta^t w(x) + \frac{1 - \beta^t}{1 - \beta} b.\end{aligned}$$

Lemma 3.1(a) follows.

(b) Since $0 < \beta < 1$, (b) follows from (a) and Assumption 3.1(ii).

To state our optimality conditions, we require some results about the discounted cost criterion. To present them we use the following notation. For each *fixed discount factor* α, $0 < \alpha < 1$, each $x \in S$, and each $\pi \in \Pi$, the discounted cost, $V_\alpha(x, \pi)$, and the corresponding discounted optimal value function, $V_\alpha^*(x)$, are as follows:

$$V_\alpha(x, \pi) := \mathrm{E}_x^\pi\left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)\right], \qquad V_\alpha^*(x) := \inf_{\pi \in \Pi} V_\alpha(x, \pi).$$

To establish the *α-discount optimality equation*, we also use the following standard continuity–compactness conditions; see, for instance, [14, p. 44], [18, p. 90], and [24, p. 15] and their references.

https://doi.org/10.1239/jap/1152413725 Published online by Cambridge University Press

Assumption 3.2. (i) *For each $x \in S$, $A(x)$ is compact.*

(ii) *For each fixed $x \in S$, $c(x, a)$ is lower semicontinuous in $a \in A(x)$, and the function $\int_S u(y)Q(\mathrm{d}y \mid x, a)$ is continuous in $a \in A(x)$ both for all bounded measurable functions u on S and for $u := w$ as in Assumption 3.1.*

Remark 3.2. Assumptions 3.2(i) and 3.2(ii) are the same as Assumption 10.2.1 of [14, p. 121]. Obviously, Assumption 3.2 holds when $A(x)$ is finite for each $x \in S$.

Lemma 3.2. *Under Assumptions 3.1 and 3.2, for each $\alpha \in (0, 1)$ the following statements hold.*

(a) $|V_\alpha(x, \pi)| \le Mw(x)/(1-\alpha) + Mb/[(1-\beta)(1-\alpha)]$ *for all $x \in S$ and $\pi \in \Pi$.*

(b) *The discounted optimal value function, $V_\alpha^*(x)$, satisfies the discounted cost optimality equation:*

$$V_\alpha^*(x) = \min_{a \in A(x)} \left\{ c(x, a) + \alpha \int_S V_\alpha^*(y) Q(\mathrm{d}y \mid x, a) \right\} \quad \text{for all } x \in S. \tag{3.2}$$

(c) *For each α, $0 < \alpha < 1$, there exists a stationary policy f_α^* (depending on α) such that $V_\alpha(x, f_\alpha^*) = V_\alpha^*(x)$ for all $x \in S$.*

Proof. By Lemma 3.1 and Assumption 3.1(ii), we have

$$\begin{aligned} |V_\alpha(x, \pi)| &\le M \sum_{t=0}^{\infty} \alpha^t \, \mathrm{E}_x^\pi [w(x_t)] \\ &\le M \sum_{t=0}^{\infty} \alpha^t \left(\beta^t w(x) + \frac{1-\beta^t}{1-\beta} b \right) \\ &\le M \sum_{t=0}^{\infty} \alpha^t \left(w(x) + \frac{b}{1-\beta} \right) \\ &= \frac{Mw(x)}{1-\alpha} + \frac{Mb}{(1-\beta)(1-\alpha)}. \end{aligned}$$

Part (a) follows.

Parts (b) and (c) follow from Theorem 8.3.6 and Remark 8.3.5 of [14, p. 46, p. 47].

To prove the existence of an average optimal stationary policy, in addition to Assumptions 3.1 and 3.2, we give a *new* condition (Assumption 3.3, below). To state this assumption, we introduce the following notation. For the function $w \ge 1$ in Assumption 3.1, we define both the weighted supremum norm, $\|u\|_w$, of a real-valued function u on S, by

$$\|u\|_w := \sup_{x \in S} w(x)^{-1} |u(x)|,$$

and the Banach space $B_w(S) := \{u : \|u\|_w < \infty\}$.

Assumption 3.3. *There exist two functions, $v_1, v_2 \in B_w(S)$, and some state, $x_0 \in S$, such that*

$$v_1(x) \le h_\alpha(x) \le v_2(x) \quad \text{for all } x \in S \text{ and } \alpha \in (0, 1),$$

where $h_\alpha(x) := V_\alpha^*(x) - V_\alpha^*(x_0)$ is, recall, the so-called relative difference *of the function $V_\alpha^*(x)$.*

https://doi.org/10.1239/jap/1152413725 Published online by Cambridge University Press

Remark 3.3. (a) Assumption 3.3 is *new* because the function $v_1(x)$ might *not* be bounded below. Moreover, Assumption 3.3 is the generalization of (SEN2) of [24, p. 132] and of Assumption 5.4.1(b) of [13, p. 86], because $h_\alpha(x)$ is assumed to be *bounded below* in those references.

(b) Assumption 3.3 holds if Assumptions 3.1 and 3.2 and the following condition (which is Assumption 10.2.2 of [14, p. 122]) are satisfied: for each $f \in F$ there exists a probability measure, μ_f, such that

$$\sup_{|u| \le w} |\mathrm{E}_x^f[u(x_t)] - \mu_f(u)| \le L\rho^t w(x) \quad \text{for all } x \in S \text{ and } t \ge 0,$$

where

$$\mu_f(u) := \int_S u(y)\mu_f(\mathrm{d}y)$$

and $L > 0$ and ρ, $0 < \rho < 1$, are constants independent of f.

To verify Assumption 3.3, we now provide some *new* sufficient conditions, and, for ease of reference, also state some existing conditions.

Lemma 3.3. *Let the function w be as in Assumption 3.1. Then, under Assumptions 3.1 and 3.2, each of the following five sets of conditions guarantees Assumption 3.3 to hold.*

(a) *For each $f \in F$, the corresponding Markov processes $\{x_t\}$ are uniformly w-*exponentially ergodic*; that is, there exists a probability measure, μ_f (depending on f), such that*

$$\sum_{t=0}^{\infty} r^t \|Q_f^t(\cdot \mid x) - \mu_f(\cdot)\|_w \le Lw(x) + b' \quad \textit{for all } x \in S, \tag{3.3}$$

where

$$\|Q_f^t(\cdot \mid x) - \mu_f(\cdot)\|_w := \sup_{|u| \le w} |\mathrm{E}_x^f[u(x_t)] - \mu_f(u)|$$

and $L > 0$, $r > 1$, and $b' \ge 0$ are constants independent of f.

(b) *$S = [0, \infty)^d$ for some integer $d \ge 1$, and the following conditions are satisfied.*

(i) *The process $\{x_t\}$ with transition law $Q(\cdot \mid x, f(x))$ is stochastically ordered (monotonic) for each $f \in F$.*

(ii) *The function w is nondecreasing and satisfies*

$$\int_S w(y)Q(\mathrm{d}y \mid x, f(x)) \le \beta w(x) + b\,\mathbf{1}_{\{0\}}(x) \quad \textit{for all } f \in F \textit{ and } x \ge 0,$$

where $\mathbf{1}_D$ denotes the indicator function of any set D and β and b are as in Assumption 3.1.

(c) *For each $f \in F$, there exist an 'atom', c_f (depending on f), in $\mathcal{B}(S)$ (such that, e.g. $Q(\cdot \mid x, f(x)) \equiv Q(\cdot \mid c_f)$ is independent of $x \in c_f$) and constants, $\delta_1, b_1 > 0$, and β_1, $0 < \beta_1 < 1$, independent of f, such that*

$$Q(c_f \mid c_f) \ge \delta_1,$$

$$\int_S w(y)Q(\mathrm{d}y \mid x, f(x)) \le \beta_1 w(x) + b_1\,\mathbf{1}_{c_f}(x) \quad \textit{for all } x \in S.$$

https://doi.org/10.1239/jap/1152413725 Published online by Cambridge University Press

(d) *For each* $f \in F$, *the transition law* $Q(\cdot \mid x, f(x))$ *has a unique invariant probability measure,* μ_f, *and, moreover, there exist a measure function,* l_f, $0 \le l_f \le 1$ *(depending on* f*), on* S, *a probability measure,* ν, *on* S, *and constants,* $\delta_2 > 0$ *and* β_2, $0 < \beta_2 < 1$, *independent of* f, *such that*

(i) $Q(B \mid x, f(x)) \ge l_f(x)\nu(B)$ *for all* $B \in \mathcal{B}(S)$ *and* $x \in S$,

(ii) $\int_S l_f(y)\nu(\mathrm{d}y) \ge \delta_2$,

(iii) $\nu(w) := \int_S w(y)\nu(\mathrm{d}y) < \infty$,

(iv) $\int_S w(y)Q(\mathrm{d}y \mid x, f(x)) \le \beta_2 w(x) + l_f(x)\nu(w)$ *for all* $x \in S$.

(e) *For each* $(x, a) \in K$, $Q(\cdot \mid x, a)$ *has a density function,* $q(x, a, y)$, *on* $K \times S$ *with respect to a measure* m, *and there exist two sets,* D_1 *and* D_2, *with* $m(D_k) > 0$, $k = 1, 2$, *and positive constants* $\delta_3 > 0$, $b_3 > 0$, *and* β_3, $0 < \beta_3 < 1$, *such that*

(i) $q(x, a, y) \ge \delta_3$ *for all* $x \in D_1$, $a \in A(x)$, *and* $y \in D_2$,

(ii) $\int_S w(y)q(x, a, y)m(\mathrm{d}y) \le \beta_3 w(x) + b_3 \mathbf{1}_{D_1}(x)$ *for all* $(x, a) \in K$,

(iii) *there exists an integer,* N, *such that* $\mathrm{P}_x^f(x_N \in D_1) \ge \delta_3$ *for all* $f \in F$ *and* $x \in \{x \in S\colon w(x) \le c\}$ *for all* $c \ge 0$.

Proof. (a) As $|c(x, a)| \le Mw(x)$ with $w(x) \ge 1$, using Assumptions 3.1 and 3.2, it follows from Lemma 3.2(c) and (3.3) that, for each $x \in S$ and α, $0 < \alpha < 1$,

$$
\begin{aligned}
|h_\alpha(x)| &= \left| \mathrm{E}_x^{f_\alpha^*}\left[\sum_{t=0}^{\infty} \alpha^t c(x_t, f_\alpha^*(x_t))\right] - \mathrm{E}_{x_0}^{f_\alpha^*}\left[\sum_{t=0}^{\infty} \alpha^t c(x_t, f_\alpha^*(x_t))\right]\right| \\
&\le \sum_{t=0}^{\infty} \alpha^t |\,\mathrm{E}_x^{f_\alpha^*}[c(x_t, f_\alpha^*(x_t))] - \mathrm{E}_{x_0}^{f_\alpha^*}[c(x_t, f_\alpha^*(x_t))]| \\
&\le \sum_{t=0}^{\infty} r^t |\,\mathrm{E}_x^{f_\alpha^*}[c(x_t, f_\alpha^*(x_t))] - \mathrm{E}_{x_0}^{f_\alpha^*}[c(x_t, f_\alpha^*(x_t))]| \\
&= M \sum_{t=0}^{\infty} r^t \left| \mathrm{E}_x^{f_\alpha^*}\left[\frac{c(x_t, f_\alpha^*(x_t))}{M}\right] - \mathrm{E}_{x_0}^{f_\alpha^*}\left[\frac{c(x_t, f_\alpha^*(x_t))}{M}\right]\right| \\
&\le ML\left(1 + w(x_0) + \frac{2b'}{L}\right) w(x) \\
&=: v_2(x),
\end{aligned}
$$

which yields Assumption 3.3 with $v_1(x) = -ML(1 + w(x_0) + 2b'/L)w(x)$.

(b) From the proof of Equation (14) of [23], for each $x \in S$, $f \in F$, and r, $1 < r \le \beta^{-1}$, we have

$$
\sum_{t=0}^{\infty} r^t \| Q_f^t(\cdot \mid x) - \mu_f(\cdot)\|_w \le \frac{2}{1 - \beta r}\left[w(x) + \frac{b}{1 - \beta}\right],
$$

which, together with (a), yields Assumption 3.3.

(c) By Theorem 2.2 of [17], we see that (c) implies (a) and, thus, yields Assumption 3.3.

https://doi.org/10.1239/jap/1152413725 Published online by Cambridge University Press

(d) By Proposition 10.2.5 of [14, p. 126], we see that (d) implies (a) and, thus, yields Assumption 3.3.

(e) By Theorem 5.1 and Lemma 5.1 of [16], we see that (e) implies (a) and, thus, yields Assumption 3.3.

Remark 3.4. (a) Conditions (a) and (b) in Lemma 3.3 are *different* from those in [2], [7, p. 188], [8], [14, p. 122], [13, p. 96], [12, p. 56], [15], and [16]. In particular, the stochastic monotonicity condition (b)(i) has been used to verify Assumption 3.3. These conditions are the generalization of ergodic conditions of [12, p. 56] and the minorant conditions of [7, p. 188].

(b) Condition (c) in Lemma 3.3 is a variant of those of Theorem 2.2 of [17], and conditions (d) and (e) in Lemma 3.3 follow from Proposition 10.2.5 of [14, p. 126] and Theorem 5.1 of [16], respectively.

4. Existence of average optimal stationary policies

In this section we provide our main results.

Theorem 4.1. *Under Assumptions 3.1, 3.2, and 3.3, the following assertions hold.*

(a) *There exist a unique constant, g^*, two functions, $h_1^*, h_2^* \in B_w(S)$, and a stationary policy, $f^* \in F$, satisfying the two average optimality inequalities*

$$\rho + h_1^*(x) \le \min_{a\in A}\left\{c(x,a) + \int_S h_1^*(y)Q(\mathrm{d}y \mid x,a)\right\} \quad \textit{for all } x \in S, \tag{4.1}$$

$$g^* + h_2^*(x) \ge \min_{a\in A(x)}\left\{c(x,a) + \int_S h_2^*(y)Q(\mathrm{d}y \mid x,a)\right\} \tag{4.2}$$

$$= c(x, f^*(x)) + \int_S h_2^*(y)Q(\mathrm{d}y \mid x, f^*(x)) \quad \textit{for all } x \in S. \tag{4.3}$$

(b) $g^* = \inf_{\pi\in\Pi} \bar{V}(x,\pi)$ *for all* $x \in S$.

(c) *Any stationary policy, $f \in F$, realizing the minimum of (4.2) is average optimal; thus, f^* in (4.3) is an average optimal stationary policy.*

Proof. (a) Let x_0 be as in Assumption 3.3, and let $\{\alpha_n\}$ be any sequence of *increasing* discount factors such that $\alpha_n \to 1$ as $n \to \infty$. By Lemma 3.2(a), $(1-\alpha_n)V_{\alpha_n}^*(x_0)$ is bounded for $n \ge 1$. Therefore, there exist a subsequence, $\{\alpha_k\}$, of $\{\alpha_n\}$ and a constant, g^*, satisfying

$$\lim_{k\to\infty}(1-\alpha_k)V_{\alpha_k}^*(x_0) = g^*, \qquad h_1^*(x) := \limsup_{k\to\infty} h_{\alpha_k}(x). \tag{4.4}$$

Since $|h_{\alpha_k}| \le |v_1| + |v_2|$, h_1^* belongs to $B_w(x)$ (by Assumption 3.3). Recalling that $h_\alpha(x) = V_\alpha^*(x) - V_\alpha^*(x_0)$, from (3.2) we have

$$(1-\alpha)V_\alpha^*(x_0) + h_\alpha(x) = \min_{a\in A(x)}\left\{c(x,a) + \alpha\int_S h_\alpha(y)Q(\mathrm{d}y \mid x,a)\right\} \quad \text{for all } x \in S,$$

which yields

$$(1-\alpha_k)V_{\alpha_k}^*(x_0) + h_{\alpha_k}(x) \le c(x,a) + \int_S \alpha_k h_{\alpha_k}(y)Q(\mathrm{d}y \mid x,a) \quad \text{for all } x \in S \text{ and } a \in A(x). \tag{4.5}$$

https://doi.org/10.1239/jap/1152413725 Published online by Cambridge University Press

By applying Lemma 8.3.7 of [14, p. 48] (an 'extension of Fatou's lemma') and letting $k \to \infty$ in (4.5), by (4.4) we have

$$g^* + h_1^*(x) \le c(x,a) + \int_S h_1^*(y) Q(\mathrm{d}y \mid x,a) \quad \text{for all } x \in S \text{ and } a \in A(x),$$

which yields

$$g^* + h_1^*(x) \le \min_{a \in A(x)} \left\{ c(x,a) + \int_S h_1^*(y) Q(\mathrm{d}y \mid x,a) \right\} \quad \text{for all } x \in S.$$

Equation (4.1) follows from this.

Now we prove (4.2). For each $x \in S$, let

$$h_2^*(x) := \liminf_{k \to \infty} h_{\alpha_k}(x) \in B_w(S),$$

whence

$$h_2^*(x) = \lim_{k \to \infty} g_{\alpha_k}(x) \quad \text{with}$$
$$g_{\alpha_k}(x) := \inf\{h_{\alpha_m}(x) \colon m \ge k\} \le h_{\alpha_k}(x) \quad \text{for all } k \ge 1.$$

Similarly, by (3.2) and $h_{\alpha_k} \ge g_{\alpha_k}$, we have

$$\begin{aligned}(1-\alpha_k) V_{\alpha_k}^*(x_0) + h_{\alpha_k}(x) &= \min_{a \in A(x)} \left\{ c(x,a) + \alpha_k \int_S h_{\alpha_k}(y) Q(\mathrm{d}y \mid x,a) \right\} \\ &\ge \min_{a \in A(x)} \left\{ c(x,a) + \alpha_k \int_S g_{\alpha_k}(y) Q(\mathrm{d}y \mid x,a) \right\}. \end{aligned} \tag{4.6}$$

Since $\alpha_{k+1} \ge \alpha_k > 0$ and $g_{\alpha_{k+1}} - g_{\alpha_1} \ge g_{\alpha_k} - g_{\alpha_1} \ge 0$, we see that the limits

$$\lim_{k \to \infty} \min_{a \in A(x)} \left\{ c(x,a) + \alpha_k \int_S (g_{\alpha_k}(y) - g_{\alpha_1}(y)) Q(\mathrm{d}y \mid x,a) \right\}$$

and

$$\lim_{k \to \infty} \min_{a \in A(x)} \left\{ c(x,a) + \alpha_k \int_S g_{\alpha_k}(y) Q(\mathrm{d}y \mid x,a) \right\} \tag{4.7}$$

exist. By (4.4) and (4.6), we then have

$$g^* + h_2^*(x) \ge \lim_{k \to \infty} \min_{a \in A(x)} \left\{ c(x,a) + \alpha_k \int_S g_{\alpha_k}(y) Q(\mathrm{d}y \mid x,a) \right\}. \tag{4.8}$$

On the other hand, for each fixed $k \ge 1$, by Assumption 3.2, there exists an $a_k(x) \in A(x)$ (depending on k and x) such that

$$\begin{aligned} &\min_{a \in A(x)} \left\{ c(x,a) + \alpha_k \int_S g_{\alpha_k}(y) Q(\mathrm{d}y \mid x,a) \right\} \\ &\quad = c(x, a_k(x)) + \alpha_k \int_S g_{\alpha_k}(y) Q(\mathrm{d}y \mid x, a_k(x)). \end{aligned} \tag{4.9}$$

https://doi.org/10.1239/jap/1152413725 Published online by Cambridge University Press

Since $A(x)$ is compact, there exists a subsequence, $\{a_{k_i}(x)\}$, of $\{a_k(x)\}$ such that the limit $\lim_{i\to\infty} a_{k_i}(x)$ exists; we denote it by $a'(x) \in A(x)$. Noting that $\|g_{\alpha_k}\|_w \le \|v_1\|_w + \|v_2\|_w$ for all $k \ge 1$, from Assumption 3.2(ii), (4.7)–(4.9), and Lemma 8.3.7 of [14, p. 48] we obtain

$$\begin{aligned} g^* + h_2^*(x) &\ge \lim_{i\to\infty} \min_{a\in A(x)} \left\{ c(x,a) + \alpha_{k_i} \int_S g_{\alpha_{k_i}}(y) Q(\mathrm{d}y \mid x, a) \right\} \\ &= \lim_{i\to\infty} \left[c(x, a_{k_i}(x)) + \alpha_{k_i} \int_S g_{\alpha_{k_i}}(y) Q(\mathrm{d}y \mid x, a_{k_i}(x)) \right] \\ &\ge c(x, a'(x)) + \int_S h_2^*(y) Q(\mathrm{d}y \mid x, a'(x)) \\ &\ge \min_{a\in A(x)} \left\{ c(x,a) + \int_S h_2^*(y) Q(\mathrm{d}y \mid x, a) \right\}. \end{aligned}$$

Equation (4.2) follows from this. Moreover, (4.2) together with the well-known 'measurable selection theorem' [14, p. 50] implies the existence of an $f^* \in F$ satisfying (4.3). Thus, the proof of part (a) is complete.

(b) For each $\pi \in \Pi$ and $x \in S$, from (4.1) we obtain

$$g^* + h_1^*(x_t) \le c(x_t, a_t) + \int_S h_1^*(y) Q(\mathrm{d}y \mid x_t, a_t) \quad \text{for all } x_t \in S,\ a_t \in A(x_t), \text{ and } t \ge 0,$$

which, together with (2.1), yields

$$g^* + \mathrm{E}_x^\pi[h_1^*(x_t)] \le \mathrm{E}_x^\pi[c(x_t, a_t)] + \mathrm{E}_x^\pi[h_1^*(x_{t+1})] \quad \text{for all } t \ge 0$$

and, thus,

$$g^* + \frac{h_1^*(x)}{N} \le \frac{\mathrm{E}_x^\pi[\sum_{t=0}^{N-1} c(x_t, a_t)]}{N} + \frac{\mathrm{E}_x^\pi[h_1^*(x_N)]}{N} \quad \text{for all } N \ge 1. \tag{4.10}$$

However, by Lemma 3.1(a) we have

$$\mathrm{E}_x^\pi[|h_1^*(x_N)|] \le \|h_1^*\|_w \left[\beta^N w(x) + \frac{1-\beta^N}{1-\beta} b \right].$$

Hence, we have $\lim_{N\to\infty} \mathrm{E}_x^\pi[h_1^*(x_N)]/N = 0$, which, together with (4.10), yields

$$g^* \le \bar{V}(x, \pi) \quad \text{for all } x \in S \text{ and } \pi \in \Pi. \tag{4.11}$$

Thus,

$$g^* \le \inf_{\pi\in\Pi} \bar{V}(x, \pi) \quad \text{for all } x \in S. \tag{4.12}$$

Similarly, by (4.3) we have

$$g^* + h_2^*(x_t) \ge c(x_t, f^*(x_t)) + \int_S h_2^*(y) Q(\mathrm{d}y \mid x_t, f^*(x_t)) \quad \text{for all } x_t \in S \text{ and } t \ge 0. \tag{4.13}$$

Then, as in the proof of (4.11), by (4.13) we have

$$g^* \ge \bar{V}(x, f^*) \quad \text{for all } x \in S. \tag{4.14}$$

By (4.12) and (4.14), we have $g^* = \bar{V}(x, f^*) = \inf_{\pi\in\Pi} \bar{V}(x, \pi)$, completing the proof of part (b).

https://doi.org/10.1239/jap/1152413725 Published online by Cambridge University Press

(c) The proof of part (c) follows obviously from the proofs of parts (a) and (b).

Remark 4.1. (a) It should be mentioned that there are *two key steps* in the 'optimality inequality approach' used, for instance, in [14], [13], [18], [25], and [24]. The first step is to obtain an inequality such as (4.12) by the Abelian theorem relating the average cost, $\bar{V}(x,\pi)$, to the discounted cost, $V_\alpha(x,\pi)$, and the other is to obtain an inequality such as (4.14) from an optimality inequality such as (4.3). To guarantee the applicability of the Abelian theorem, the costs have to be *nonnegative*. Thus, the Abelian theorem used in the optimality inequality approach in the previous literature is *not* applicable to our case, because the costs in our model may have *neither upper nor lower bounds*. Therefore, the approach provided in this paper may be regarded as a modification of the optimality inequality approach taken in previous works cited.

(b) When the state space S is *denumerable*, under Assumptions 3.1, 3.2, and 3.3 the standard *diagonalization argument* serves to show the existence of a sequence, $\{h_{\alpha_k}(x)\}$, such that the limit $h^*(x) := \lim_{k\to\infty} h_{\alpha_k}(x)$ exists for all $x \in S$. Hence, $h_1^*(x) = h_2^*(x) = h^*(x)$ for all $x \in S$, the inequalities (4.1) and (4.2) thus coincide, and the AOE is obtained. Moreover for a denumerable state space, under suitable conditions some stronger results have been obtained; see, for instance, [5] for the existence of a Blackwell optimal policy and [1] for a condition sufficient and necessary for an optimal policy.

(c) To establish the AOE in Borel spaces, in addition to our Assumptions 3.1, 3.2 and 3.3, an *additional* condition is required; see, for instance, the irreducibility condition (e.g. Assumption 10.3.5 of [14, p. 130].) On the other hand, under the slightly stronger conditions of [15] and [16], not only can the AOE be established using Lemmas 5.1, 6.1, and 6.3 of [15] (or by Theorem 10.3.6 of [14, p. 130]), but the existence of average and Blackwell optimal policies can also be shown.

5. Examples

In this section we first illustrate our assumptions with a generalized inventory system and then apply our results to a controlled queueing model.

Example 5.1. (*A generalized inventory system.*) Consider a control system of the form

$$x_{t+1} = (x_t + a_t\eta_t - \xi_t)^+, \qquad t = 0, 1, \ldots, \tag{5.1}$$

with a state space $S := [0, \infty)$. This model can in fact have several interesting interpretations, such as, as a random-release dam model or a single-server queueing system (of general type GI/GI/1) with controllable service rates. Here, we interpret (5.1) as a *generalized inventory system*. Thus, x_t and η_t respectively denote the stock level and amount of 'base product' ordered (and immediately supplied) at the beginning of period t, while ξ_t denotes the demand during period t. The control variable, a_t, denotes the reciprocal of the amount of base product ordered at the beginning of period t. We denote by $c(x,a)$ an associated cost function for this system.

For the average optimality of system (5.1), we consider the following hypotheses.

Assumption 5.1. (On Example 5.1.) (i) *For each $x \in S$, the action set $A(x)$ is a compact subset of an interval $(0,\theta]$, for some finite $\theta > 0$.*

(ii) *$\{\eta_t\}$ and $\{\xi_t\}$ are independent sequences of independent, identically distributed random variables.*

https://doi.org/10.1239/jap/1152413725 Published online by Cambridge University Press

(iii) η_0 *takes finite values* $b_i \geq 0$, $1 \leq i \leq N < \infty$, *and has probability distribution* $p_i := \mathrm{P}(\eta_0 = b_i)$ *with* $\sum_{i=1}^{N} p_i = 1$, *while* ξ_0 *has a continuous and bounded density function* q. *In particular, system (5.1) is the well-known inventory system of a random-release dam model when* $\mathrm{P}(\eta_0 = 1) = 1$.

(iv) $\mathrm{E}[z_\theta] < 0$ *and* $\psi_\theta(\bar{r}) < \infty$ *for some* $\bar{r} > 0$, *where* $\mathrm{E}[z_a]$ *and* $\psi_a(r) := \mathrm{E}[\mathrm{e}^{rz_a}]$ *respectively denote the mean and a moment generating function of the random variable* $z_a := a\eta_0 - \xi_0$ *with* $a \in (0, \theta]$. *(Thus,* $\psi_\theta(0) = 1$ *and* $\psi'_\theta(0) = \mathrm{E}[z_\theta] < 0$, *and, so, there exists a constant,* ρ, $0 < \rho < \bar{r}$, *such that* $\psi_\theta(\rho) < 1$.*)*

(v) *There exists a constant,* $M > 0$, *such that* $|c(x, a)| \leq M\mathrm{e}^{\rho x}$ *for all* $x \in S$ *and* $a \in A(x)$, *with* ρ *as in part (iv).*

(vi) $\psi_a(\rho)$ *is continuous in* $a \in A(x)$ *for each* $x \in S$, *and* $c(x, a)$ *is lower semicontinuous in* $a \in A(x)$.

We now define the weight function $w(x) := \mathrm{e}^{\rho x}$ for all $x \in S$, and proceed to verify Assumptions 3.1, 3.2, and 3.3. By Assumptions 5.1(vi) and 5.1(v) and the description of the model, we find that Assumptions 3.1(ii) and 3.2(i) are automatically satisfied. Thus, it only remains to verify Assumptions 3.2(ii), 3.1(i), and 3.3.

Verification of Assumption 3.2(ii). For each $a \in A(x)$, we have $z_a = a\eta_0 - \xi_0 \leq z_\theta$ (as $0 < a \leq \theta$ for all $a \in A$). Thus, by Assumptions 5.1(ii) and 5.1(iii), we can derive the distribution function $G(a, y)$ of z_a, as follows:

$$G(a, y) := \mathrm{P}(z_a \leq y) = \sum_{i=1}^{N} \mathrm{P}(ab_i - \xi_0 \leq y)p_i = \sum_{i=1}^{N} \int_{ab_i - y}^{\infty} q(v)p_i \,\mathrm{d}v. \tag{5.2}$$

Hence, from Assumption 5.1(iii) and (5.2) we see that the density function of z_a,

$$g(a, y) = \sum_{i=1}^{N} q(ab_i - y)p_i, \tag{5.3}$$

is a bounded, continuous function of $a \in A(x)$, and it follows from (5.2) that $G(a, y)$ is also bounded and continuous in $a \in A(x)$. Thus, for each measurable bounded function u on S, by (5.1) and (5.3) we see that

$$\int_S u(y)Q(\mathrm{d}y \mid x, a) = u(0)G(a, -x) + \int_{-x}^{\infty} u(x + y)g(a, y) \,\mathrm{d}y \tag{5.4}$$

is also bounded and continuous in $a \in A(x)$. Moreover, by replacing u in (5.4) by the weight function w above, and noting that $z_a \leq z_\theta$, we obtain

$$\begin{aligned}
\int_S w(y)Q(\mathrm{d}y \mid x, a) &= G(a, -x) + \int_{-x}^{\infty} w(x + y)g(a, y) \,\mathrm{d}y \\
&= G(a, -x) + w(x) \int_{-x}^{\infty} \mathrm{e}^{\rho y} g(a, y) \,\mathrm{d}y \\
&= G(a, -x) + w(x)\left[\psi_a(\rho) - \int_{-\infty}^{-x} \mathrm{e}^{\rho y} g(a, y) \,\mathrm{d}y\right] \qquad (5.5) \\
&\leq \psi_\theta(\rho)w(x) + G(a, -x). \qquad (5.6)
\end{aligned}$$

https://doi.org/10.1239/jap/1152413725 Published online by Cambridge University Press

Since $g(a, y)$ is bounded and continuous in $a \in A(x)$, by Assumption 5.1(vi) and (5.5) we see that $\int_S w(y)Q(\mathrm{d}y \mid x, a)$ is continuous in $a \in A(x)$, and Assumption 3.2(ii) thus follows.

Verification of Assumption 3.1(i). By (5.6) we can verify Assumption 3.1(i) with $\beta \equiv \psi_\theta(\rho)$ and $b \equiv 1$.

Verification of Assumption 3.3. We use Lemma 3.3(d) to verify Assumption 3.3. To do so, let $f \in F$, let ν be the Dirac measure at $x = 0$, and define

$$\begin{aligned} l_a(x) &:= G(a, -x) \qquad \text{for all } x \in S \text{ and } a \in A(x), \\ l_f(x) &:= G(f(x), -x) \quad \text{for all } x \in S. \end{aligned}$$

Then, as $0 < f(x) \le \theta$ for all $x \in S$, we have

$$l_f(x) = G(f(x), -x) = \mathrm{P}(f(x)\eta_0 - \xi_0 \le -x) \ge \mathrm{P}(\theta\eta_0 - \xi_0 \le -x) = G(\theta, -x),$$

which, together with (5.4) and (5.6), yields

$$\begin{aligned} Q(B \mid x, f(x)) &\ge l_f(x)\nu(B) \quad \text{for all } B \in \mathcal{B}(S) \text{ and } x \in S, \\ \int_S w(y)Q(\mathrm{d}y \mid x, f(x)) &\le \psi_\theta(\rho)w(x) + l_f(x) \qquad \text{for all } x \in S, \end{aligned}$$

from which parts (d)(i) and (d)(iv) of Lemma 3.3 follow.

Moreover, by the definitions of ν and l_f, we have

$$\begin{aligned} \int_S l_f(y)\nu(\mathrm{d}y) = l_f(0) \ge G(\theta, 0) > 0 \quad \text{for all } f \in F, \\ \nu(w) := \int_S w(y)\,\mathrm{d}\nu(\mathrm{d}y) = 1, \end{aligned}$$

from which parts (d)(ii) and (d)(iii) of Lemma 3.3 follow. Hence, all conditions in Lemma 3.3(d) have been verified.

Exactly as in Lemma 10.9.4 of [14, p. 159], we can also verify that $Q(\cdot \mid x, f(x))$ (for each fixed $f \in F$) has a unique invariant probability measure, μ_f. Thus, Assumption 3.3 follows from Lemma 3.3.

In summary, we have the following result.

Proposition 5.1. *Under Assumption 5.1, the generalized inventory system above satisfies Assumptions 3.1, 3.2, and 3.3. Therefore (by Theorem 4.1), there exists an average optimal stationary policy.*

Remark 5.1. (a) According to the discussions above, we see that for Example 5.1 all conditions in this paper are satisfied. It should be noted that in Example 5.1 the state space is not denumerable and that we can easily give a cost function (for this example) that might have *neither upper nor lower bounds*. Therefore, the earlier conditions in [1], [2], [3], [4], [7, p. 184], [5], [14, p. 128], [13, p. 46], [12, p. 7], [18, p. 18], [19], [22], [20], [25], and [24, p. 15] fail to hold. This is because the state spaces considered in the previous literature are all denumerable, except in [2], [7], [14], [13], [12], [19], and [22], where the cost functions are assumed to be bounded below.

(b) Although the conditions in this paper are slightly weaker than those in [15] and [16], a straightforward calculation together with the Radon–Nikodým theorem can show that the

results of [15] and [16] also apply to Example 5.1. Moreover, it should be mentioned that it is very interesting and difficult to find a real model for which all conditions in this paper are fulfilled, but which does not satisfy the assumptions of [15] and [16].

We now apply our results to a queueing system.

Example 5.2. (*A controlled queueing system.*) Consider a controlled queueing system in which the state variable denotes the number of customers in the system. When there is more than one customer in the system, we suppose that an arriving customer will be rejected from the system with probability $p_1 > 0$ and admitted to the system with probability $p_2 := 1 - p_1$. The arrival rate is assumed to be a *fixed* constant, $\lambda > 0$, and the service rates a are assumed to be controlled by a decision-maker. Here, we interpret the service rates as the *control actions*. When the system's state is $x \in S := \{0, 1, \dots\}$, the decision-maker takes an action a from a given set $A(x) \equiv [\mu_1, \mu_2]$ with $\mu_2 > \mu_1 > 0$, which increases or decreases the service rates given by (5.7)–(5.9), below. The action incurs a cost $\bar{c}(x, a)$. In addition, the decision-maker obtains a reward px during the time in which the system remains in state x ($p > 0$ denotes the unit reward produced by a customer).

We now formulate this system as a discrete-time Markov decision process. The corresponding transition law, $Q(\cdot \mid x, a)$, and cost function, $c(x, a)$, are given as follows. When there is at most one customer in the system, it is natural to assume that no control of the system is necessary. Thus, for each $a \in A(x)$ with $x = 0, 1$, we have

$$Q(0 \mid 0, a) = \frac{\mu_2}{\lambda + \mu_2}, \qquad Q(1 \mid 0, a) = \frac{\lambda}{\lambda + \mu_2}, \tag{5.7}$$

$$Q(0 \mid 1, a) = \frac{\mu_2}{\lambda + \mu_2}, \qquad Q(1 \mid 1, a) = \frac{p_1\lambda}{\lambda + \mu_2}, \qquad Q(2 \mid 1, a) = \frac{p_2\lambda}{\lambda + \mu_2}. \tag{5.8}$$

Moreover, for each $x \geq 2$ and $a \in A(x)$,

$$Q(y \mid x, a) := \begin{cases} \dfrac{a}{\lambda + \mu_2} & \text{if } y = x - 2, \\ \dfrac{\mu_2 - a}{\lambda + \mu_2} & \text{if } y = x - 1, \\ \dfrac{p_1\lambda}{\lambda + \mu_2} & \text{if } y = x, \\ \dfrac{p_2\lambda}{\lambda + \mu_2} & \text{if } y = x + 1, \\ 0 & \text{otherwise,} \end{cases} \tag{5.9}$$

$$c(x, a) := \bar{c}(x, a) - px \quad \text{for } (x, a) \in K := \{(x, a) \colon x \in S,\ a \in A(x)\}. \tag{5.10}$$

We aim to find conditions that ensure the existence of an average optimal stationary policy. To do this, we consider the following assumptions.

Assumption 5.2. *The parameter λ is such that* $(\mathrm{e} - 1)\mu_2 > p_2\lambda\mathrm{e}^2$.

Assumption 5.3. *The function $\bar{c}(x, a)$ is continuous in $a \in A(x)$ for each fixed $x \in S$, and such that $|\bar{c}(x, a)| \leq L\mathrm{e}^x$ for all $(x, a) \in K$ and some constant, $L > 0$.*

https://doi.org/10.1239/jap/1152413725 Published online by Cambridge University Press

Under these conditions, we have the following result.

Proposition 5.2. *Under Assumptions 5.2 and 5.3, the controlled queueing system above satisfies Assumptions 3.1, 3.2, and 3.3. Therefore (by Theorem 4.1), there exists an average optimal stationary policy.*

Proof. We shall first verify Assumption 3.1. Let $\rho = [\mu_2 + p_1\lambda e + p_2\lambda e^2]/[e(\lambda + \mu_2)]$ and let $w(x) = e^x$ for all $x \in S$. By Assumption 5.2 we see that $0 < \rho < 1$. Then, combining (5.7) and (5.8), we have

$$\sum_{y\in S} Q(y \mid 0, a)w(y) = \frac{\mu_2}{\lambda+\mu_2} + \frac{\lambda e}{\lambda+\mu_2} \leq \rho w(0) + \frac{\mu_2+\lambda e}{\lambda+\mu_2}, \tag{5.11}$$

$$\sum_{y\in S} Q(y \mid 1, a)w(y) = \frac{\mu_2}{\lambda+\mu_2} + \frac{p_1\lambda}{\lambda+\mu_2}e + \frac{p_2\lambda}{\lambda+\mu_2}e^2 = \rho w(1). \tag{5.12}$$

Moreover, for each $x \geq 2$ and $a \in A(x)$, from (5.9) it follows that

$$\begin{aligned}\sum_{y\in S} Q(y \mid x, a)w(y) &= \frac{a}{\lambda+\mu_2}e^{x-2} + \frac{\mu_2-a}{\lambda+\mu_2}e^{x-1} + \frac{p_1\lambda}{\lambda+\mu_2}e^x + \frac{p_2\lambda}{\lambda+\mu_2}e^{x+1}\\ &= \frac{a + e(\mu_2-a) + p_1\lambda e^2 + p_2\lambda e^3}{e^2(\lambda+\mu_2)}w(x)\\ &\leq \frac{e\mu_2 + \mu_1(1-e) + p_1\lambda e^2 + p_2\lambda e^3}{e^2(\lambda+\mu_2)}w(x)\\ &\leq \rho w(x).\end{aligned} \tag{5.13}$$

Thus, for each $x \in S$ and $a \in A(x)$, by (5.11)–(5.13) we have

$$\sum_{y\in S} Q(y \mid x, a)w(y) \leq \rho w(x) + \frac{\mu_2+\lambda e}{\lambda+\mu_2}\mathbf{1}_{\{0\}}(x) \leq \rho w(x) + \frac{\mu_2+\lambda e}{\lambda+\mu_2}, \tag{5.14}$$

which gives Assumption 3.1(i) with $b = (\mu_2 + \lambda e)/(\lambda + \mu_2)$ and $\beta = \rho < 1$ defined as above. On the other hand, since $e^x \geq x$, from (5.10) and Assumption 5.3 we have

$$|c(x, a)| \leq (p + L)w(x) \quad \text{for all } x \in S,$$

which verifies Assumption 3.1(ii) with $M := p + L$. Assumption 3.1 is thus satisfied.

By (5.7)–(5.9), the model's description, and Assumption 5.3, Assumption 3.2 is obviously satisfied.

Finally, we verify Assumption 3.3. By (5.7)–(5.9), for each fixed $f \in F$ we have

$$\sum_{y\geq k} Q(y \mid x, f(x)) \leq \sum_{y\geq k} Q(y \mid x', f(x')) \quad \text{for all } x, x', k \in S \text{ and } x < x',$$

https://doi.org/10.1239/jap/1152413725 Published online by Cambridge University Press

which, together with Theorem 7.4.1 of [21, p. 298], implies that the corresponding Markov process, $\{x_t\}$, is stochastically monotone (for each $f \in F$). Thus, by (5.14) and Lemma 3.3(b) we see that Assumption 3.3 is satisfied.

6. Concluding remarks

In the previous sections we have studied the average optimality problem for discrete-time Markov decision processes in Borel spaces. Under suitable assumptions we have shown the existence of an average optimal stationary policy. The approach developed in this paper is different from the optimality inequality approach widely used in the previous literature, and may be regarded as a modification thereof. We believe that our formulation and approach can be used to analyse other important problems, such as that of *stochastic games* and *average sample path optimality* for discrete-time Markov processes in Borel spaces.

Acknowledgement

The authors are greatly indebted to the anonymous referee for many valuable comments and suggestions that have improved the presentation.

References

[1] Altman, E., Hordijk, A. and Spieksma, F. M. (1979). Contraction conditions for average and α-discount optimality in countable state Markov games with unbounded rewards. *Math. Operat. Res.* **22,** 588–618.
[2] Arapostathis, A. *et al.* (1993). Discrete-time controlled Markov processes with average cost criterion: a survey. *SIAM J. Control Optimization* **31,** 282–344.
[3] Borkar, V. S. (1989). Control of Markov chains with long-run average cost criterion: the dynamic programming equations. *SIAM J. Control Optimization* **27,** 642–657.
[4] Cavazos-Cadena, R. and Fernández-Gaucherand, E. (1996). Denumerable controlled Markov chains with strong average optimality criterion: bounded and unbounded costs. *Math. Meth. Operat. Res.* **43,** 281–300.
[5] Dekker, R. and Hordijk, A. (1988). Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards. *Math. Operat. Res.* **13,** 395–420.
[6] Derman, C. (1970). *Finite State Markovian Decision Processes*. Academic Press, New York.
[7] Dynkin, E. B. and Yushkevich, A. A. (1979). *Controlled Markov Processes*. Springer, New York.
[8] Gordienko, E. and Hernández-Lerma, O. (1995). Average cost Markov control processes with weighted norms: existence of canonical policies. *Appl. Math.* **23,** 199–218.
[9] Guo, X. P. and Shi, P. (2001). Limiting average criteria for nonstationary Markov decision processes. *SIAM J. Optimization* **11,** 1037–1053.
[10] Guo, X. P., Liu, J. Y. and Liu, K. (2000). Nonhomogeneous Markov decision processes with Borel state space—the average criterion with nonuniformly bounded rewards. *Math. Operat. Res.* **25,** 667–678.
[11] Guo, X. P., Shi, P. and Zhu, W. P. (2001). Strong average optimality for controlled nonhomogeneous Markov chains. *Stoch. Anal. Appl.* **19,** 115–134.
[12] Hernández-Lerma, O. (1989). *Adaptive Markov Control Processes*. Springer, New York.
[13] Hernández-Lerma, O. and Lasserre, J. B. (1996). *Discrete-Time Markov Control Processes. Basic Optimality Criteria*. Springer, New York.
[14] Hernández-Lerma, O. and Lasserre, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes*. Springer, New York.
[15] Hordijk, A. and Yushkevich, A. A. (1999). Blackwell optimality in the class of stationary policies in Markov decision chains with a Borel state space and unbounded rewards. *Math. Meth. Operat. Res.* **49,** 1–39.
[16] Hordijk, A. and Yushkevich, A. A. (1999). Blackwell optimality in the class of all policies in Markov decision chains with a Borel state space and unbounded rewards. *Math. Meth. Operat. Res.* **50,** 421–448.
[17] Meyn, S. P. and Tweedie, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Prob.* **4,** 981–1011.
[18] Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York.
[19] Ritt, R. K. and Sennott, L. I. (1992). Optimal stationary policies in general state space Markov decision chains with finite action sets. *Math. Operat. Res.* **17,** 901–909.

[20] Robinson, D. R. (1976). Markov decision chains with unbounded costs and applications to the control of queues. *Adv. Appl. Prob.* **8,** 159–176.
[21] Rolski, T., Schmidli, H., Schmidli, V. and Teugels, J. (1998). *Stochastic Processes for Insurance and Finance*. John Wiley, Chichester.
[22] Ross, S. M. (1968). Arbitrary state Markovian decision processes. *Ann. Math. Statist.* **39,** 2118–2122.
[23] Scott, D. J. and Tweedie, R. L. (1996). Explicit rates of convergence of stochastically ordered Markov chains. In *Athens Conference on Applied Probability and Time Series*, Vol. 1, *Applied Probability* (Lecture Notes Statist. **114**), eds C. C. Heyde *et al.*, Springer, Berlin, pp. 176–191.
[24] Sennott, L. I. (1999). *Stochastic Dynamic Programming and the Control of Queueing Systems*. John Wiley, New York.
[25] Sennott, L. I. (2002). Average reward optimization theory for denumerable state spaces. In *Handbook of Markov Decision Processes* (Internat. Ser. Operat. Res. Manag. Sci. **40**), eds E. A. Feinberg and A. Shwartz, Kluwer, Boston, MA, pp. 153–172.