Kybernetika

Xiaolong Zou; Xianping Guo

Another set of verifiable conditions for average Markov decision processes with Borel spaces

Kybernetika, Vol. 51 (2015), No. 2, 276–292

Persistent URL: `http://dml.cz/dmlcz/144298`

Terms of use:

© Institute of Information Theory and Automation AS CR, 2015

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* `http://dml.cz`

ANOTHER SET OF VERIFIABLE CONDITIONS FOR AVERAGE MARKOV DECISION PROCESSES WITH BOREL SPACES

Xiaolong Zou and Xianping Guo

In this paper we give a *new* set of verifiable conditions for the existence of average optimal stationary policies in discrete-time Markov decision processes with Borel spaces and *unbounded* reward/cost functions. More precisely, we provide another set of conditions, which only consists of a Lyapunov-type condition and the common continuity-compactness conditions. These conditions are imposed on the *primitive data* of the model of Markov decision processes and thus easy to verify. We also give two examples for which all our conditions are satisfied, but some of conditions in the related literature fail to hold.

Keywords: discrete-time Markov decision processes, average reward criterion, optimal stationary policy, Lyapunov-type condition, unbounded reward/cost function

Classification: 90C40, 93E20

1. INTRODUCTION

The *average reward criterion* is one of the most popular performance criteria in discrete-time Markov decision processes (MDP). Much work has been done on this topic and we give a brief survey here. According to the state space, all of the existing works can be roughly classified into three classes: finite state space, countably infinite state space, and general state space. When the state space and the action space are both finite, the existence of an average optimal stationary policy is indeed guaranteed, see [3, pp. 165–176], [12, p. 450], for instance. However, when the state space is countably infinite, counterexamples have been constructed in [13, 14] to show that there may not exist an average optimal stationary policy. Therefore, many works have been devoted to the study on conditions which guarantee the existence of an average optimal stationary policy. For the case of a denumerable state space, see [1, pp. 298–315], [12, pp. 414–416], [14, pp. 132–157], for instance; and see [1, 5, 6], [7, pp. 86–100], [8, pp. 128–131] for the case of Borel spaces. This paper further studies the average reward criterion for discrete-time MDP in Borel spaces.

As far as we know, there are three main approaches to study the existence of an average optimal stationary policy for discrete-time MDP in Borel spaces; see [6], and [15]. The first one is the well-known “average optimality equation approach”, see [7,

DOI: 10.14736/kyb-2015-2-0276

p. 97], [8, p. 130], for instance. By analyzing the solutions of the optimality equation, the existence of an average optimal stationary policy is derived. To establish the average optimality equation, some strong conditions are needed, such as the equi-continuity of the relative difference, $h_\alpha(x)$, of the discounted optimal value function in [7, p. 96], and the uniformly w-geometric ergodicity condition together with the irreducibility of the corresponding Markov processes for each deterministic stationary policy in [8, p. 130]. Furthermore, for the verification of the uniformly w-geometric ergodicity condition, a set of the Lyapunov-type condition and the minorant condition is given in [3, 4] and [8, p. 126] respectively. The second one is the "optimality inequality approach", which needs the usage of the Abelian theorem relating the average reward criterion to the discount reward criterion, and thus it is designed for the case of nonnegative costs only. Readers may refer to [7, p. 80] where the relative difference $h_\alpha(x)$ is required to be bounded from below, in both the state x and the discounted factor α. The last one is the so-called "optimality two-inequality approach", which is developed by [6]. This approach allows a much weaker condition that the relative difference $h_\alpha(x)$ is assumed to be ω-bounded (see Assumption 3.3 in [6]), and so $h_\alpha(x)$ can be *unbounded (in states) from above and from below.* This assumption is stated as Condition C in this paper. It is shown in [6] that the uniformly w-geometric ergodicity condition in [8, p. 130] implies Condition C. For the verification of the uniformly w-geometric ergodicity condition, to the best of our knowledge, there are two sets of verifiable conditions (i. e. ones that are imposed on the *primitive data* of the model of MDP). Roughly, one of the two sets consists of a Lyapunov-type condition and the minorant condition in [4] and [8, p. 126], and another set with the Lyapunov-type and stochastic monotonicity conditions in [6].

This paper aims to give new and verifiable conditions for the existence of an average optimal stationary policy (i. e. another set of conditions for the verification of Condition C above). Inspired by the arguments in [9], we decompose $h_\alpha(x)$ into two parts: one being the expected discounted reward/cost till the first time the chain starting from any state x enters the state z (denote this stopping time by τ_z), and the other being the expected discounted reward/cost times the mean of $(\alpha^{\tau_z} - 1)$. We use a Lyapunov-type condition (i. e. the inequality (4.1) in this paper) to derive new estimates for the two parts. Basing the new estimates, we give a *new* set of conditions that also ensures the existence of an average optimal stationary policy; see Assumptions 1, 2 and Theorem 4.1 below. Our conditions consist of a so-called Lyapunov-type condition and the standard continuity-compactness conditions, which are imposed on the *primitive data* of the model of MDP and thus more easy to verify in certain situations. Unlike the verifiable conditions in [6] and in [8, p. 126], our new conditions don't need to calculate any bound for exponential convergence rates of the Markov chains governed by different deterministic stationary polices as [6], [8, p. 126] do. Thus, our conditions show that for the verification of Condition C both of the stochastic monotonicity in [6] and the minorant condition in [8, p. 126] can be dropped. To further show the advantage of our new conditions, we give two examples. One of the examples is on applications of our results to admission control problems of queueing systems in [14], and the other one is to show the difference between our conditions and those in the previous literature. It should be noted that the two examples satisfy our new conditions, whereas some of conditions in [3, 4, 6, 8] fail to hold for the second example; see Remarks 4.1 and 5.2, for details.

The rest of this paper is organized as follows. In Section 2, we define the model of MDP and state the optimality problem we are concerned with. In Section 3, we recall some existing conditions for the existence of an average optimal stationary policy. In Section 4, we give our new conditions, under which we prove the existence of an average optimal stationary policy. The last section illustrates some applications of the results with two examples.

2. THE CONTROL MODELS

In this section we introduce the control model of discrete-time MDP, state the optimality problems we are dealing with.

The model of MDP is a set of primitive data (X, A, K, Q, c) with the following meaning:

- X is the Borel state space (a Borel subset of a complete separable metric space), endowed with the Borel σ-algebra $\mathcal{B}(X)$. An element in X is denoted by $x \in X$;

- A is the Borel action space, endowed with the Borel σ-algebra $\mathcal{B}(A)$. An element in A is denoted by $a \in A$;

- The set $A(x)$ in $\mathcal{B}(A)$ denotes the set of all available actions at $x \in X$. And the set of all possible state-action combinations is denoted by K, that is $K := \{(x, a) : x \in X, a \in A(x)\}$. Moreover, K is assumed to be a Borel subset of $X \times A$;

- $Q(\cdot|x, a)$ is a stochastic kernel on X given $(x, a) \in K$, that is, given $x \in X$ and $a \in A(x)$, $Q(\cdot|x, a)$ is a probability measure on $\mathcal{B}(X)$, while given a set $B \in \mathcal{B}(X)$, $Q(B|\cdot,\cdot)$ is a $\mathcal{B}(K)$-measurable function. In what follows, when we say a measurable function we mean it's Borel measurable. $Q(B|x, a)$ gives the probability that the next state falls in B if the current state is x and action a is taken.

- $c(x, a)$: $K \to \mathbb{R}$ is assumed to be a real-valued and measurable function, and it gives the one-stage cost of a system if the current state is x and action a is taken. *(Since $c(x, a)$ is allowed to take positive and negative values in this paper, it can be interpreted as the reward other than the "cost".)*

A discrete-time MDP evolves as follows. A decision maker observes a system state at each time n. Whenever the system state is at $x \in X$, he/she chooses an action a from $A(x)$. As a consequence, he/she incurs a cost $c(x, a)$ that depends on x and a, and at the same time the system moves from the state x to a new state according to the distribution of the transition function $Q(\cdot|x, a)$. At the new state, the decision maker chooses a suitable action, and so on. In order to minimize his/her cost (or maximize the reward, when $c(x, a)$ take negative values), the decision maker aims to find a formalism of choosing the actions to be taken, i. e. a policy in MDP.

To formally define a policy, we introduce some notation: for each $n \geq 0$, let H_n denote the family of admissible histories up to time n, that is, $H_0 := X$, and $H_n := K \times H_{n-1}$ for each $n \geq 1$.

Definition 2.1. A randomized history-dependent policy is a sequence $\pi := (\pi_n, n \geq 0)$ of stochastic kernels π_n on A given H_n that satisfy

$$\pi_n(A(x)|h_n) = 1 \text{ for } \textit{all } h_n = (x_0, a_0, \ldots, x_{n-1}, a_{n-1}, x) \in H_n, \ n \geq 0.$$

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

$$\pi_n(\{f(x)\}|h_n) = \pi_n(\{f(x)\}|x) = 1 \text{ for all } h_n = (x_0, a_0, \ldots, x_{n-1}, a_{n-1}, x) \in H_n, \ n \geq 0.$$

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* X with $f(x) \in A(x)$ for all $x \in X$. Obviously, $F \subset \Pi$.

For each $x \in X$ and $\pi \in \Pi$, by the theorem of C. Ionescu Tulcea (see [7, p. 178] for instance), there exist a unique probability measure space $(\Omega, \mathscr{F}, P_x^\pi)$ and discrete-time stochastic processes $\{x_n\}$ and $\{a_n\}$, defined on Ω, such that, for each $B \in \mathcal{B}(X)$ and $n \geq 1$,

$$P_x^\pi(x_{n+1} \in B|h_n, a_n) = Q(B|x_n, a_n) \tag{2.1}$$

for $h_n = (x_0, a_0, \ldots, x_{n-1}, a_{n-1}, x_n) \in H_n$, where x_n and a_n denote the state and action variables at time $n \geq 1$, 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 $\{x_n\}$ is a Markov process with the transition law $Q_f(\cdot|x){:=}Q(\cdot|x, f(x))$.

We now state the average optimality problem we are concerned with. The total expected n-stage cost when using the policy π, given the initial state $x_0 = x$, is defined by

$$J_n(x, \pi) := E_x^\pi \left[\sum_{t=0}^{n-1} c(x_t, a_t) \right]$$

and the long run expected average cost is defined as follows:

$$J(x, \pi) := \limsup_{n \longrightarrow \infty} J_n(x, \pi)/n. \tag{2.2}$$

Definition 2.2. A policy π^* is called average optimal if

$$J(x, \pi^*) = \inf_{\pi \in \Pi} J(x, \pi) =: J^*(x) \text{ for all } x \in X, \tag{2.3}$$

and $J^*(x)$ is called the average cost value function.

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

3. EXISTING OPTIMALITY CONDITIONS AND RESULTS

In this section, we introduce the general optimality conditions (i. e. Conditions A, B and C below) given in [6], which are sufficient to guarantee the existence of an average optimal stationary policy.

Firstly, we state the so-called "expected growth condition" below (i. e. Assumption 3.1 in [6]) for the finiteness of the expected average cost $J(x, \pi)$.

Condition A. (i) *There exist positive constants $\beta < 1$ and b, and a measurable function $\omega \geq 1$ on X such that*

$$\int_X Q(\mathrm{d}y|x,a)\omega(y) \leq \beta\omega(x) + b \quad \textit{for all } x \in X,\ a \in A(x). \tag{3.1}$$

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

The following conditions are the standard continuity-compactness conditions, which are commonly used in MDP, see [8, p. 44] or Assumption 3.2 in [6] for instance.

Condition B. (i) *For each $x \in X$, $A(x)$ is compact.*
(ii) *For each fixed $x \in X$, $c(x,a)$ is lower semi-continuous in $a \in A(x)$, and the function $\int_X u(y)Q(\mathrm{d}y|x,a)$ is continuous in $a \in A(x)$ for all bounded measurable functions u on X and for $u := \omega$ defined as in Condition A.*

In addition to the above two conditions A and B, a key condition (i. e. Condition C below or Assumption 3.3 in [6]) is needed in ensuring the existence of an average optimal stationary policy. To state it, we have to introduce some notation such as the discounted cost $V_\alpha(x,\pi)$.

For a fixed discount factor $\alpha \in (0,1)$, $x \in X$, and $\pi \in \Pi$, the *discounted cost* $V_\alpha(x,\pi)$, and the corresponding *discounted optimal value function* $V_\alpha^*(x)$, are defined as follows:

$$V_\alpha(x,\pi) := E_x^\pi \left[\sum_{n=0}^{\infty} \alpha^n c(x_n,a_n)\right],\ V_\alpha^*(x) := \inf_{\pi\in\Pi} V_\alpha(x,\pi).$$

The *relative difference* of the discounted optimal value function $V_\alpha^*(x)$ is defined by $h_\alpha(x) := V_\alpha^*(x) - V_\alpha^*(z)$, where z is some fixed state.

Furthermore, for any measurable function $\omega(x) \geq 1$ on X, we define both the w-weighted norm, $\|u\|_\omega$, of a real-valued function u on X, by

$$\|u\|_\omega := \sup_{x\in X} \omega(x)^{-1}|u(x)|$$

and the Banach space of all ω-bounded measurable functions $\mathcal{B}_\omega(X) := \{u \in \mathcal{B}(X) : \|u\|_\omega < \infty\}$.

We are ready to state Condition C now (i. e. Assumption 3.3 in [6]).

Condition C. *There exist two functions, v_1, $v_2 \in \mathcal{B}_\omega(X)$, and some state, $z \in X$, such that*

$$v_1(x) \leq h_\alpha(x) \leq v_2(x) \ \textit{for all } x \in X \textit{ and } \alpha \in (0,1).$$

Under the above conditions, we have the following results, which are taken from Theorem 4.1 in [6].

Theorem 3.1. Under Conditions A, B, C, the following assertions hold:

(a) There exist a constant g^*, two functions $u_1^*,\ u_2^* \in \mathcal{B}_\omega(X)$, and a stationary policy $f^* \in F$, satisfying the following two optimality inequalities

$$g^* + u_1^*(x) \quad \le \inf_{a\in A(x)} \left\{ c(x,a) + \int_X u_1^*(y) Q(\mathrm{d}y|x,a) \right\} \text{ for all } x \in X, \tag{3.2}$$

$$g^* + u_2^*(x) \quad \ge \inf_{a\in A(x)} \left\{ c(x,a) + \int_X u_2^*(y) Q(\mathrm{d}y|x,a) \right\} \tag{3.3}$$

$$= c(x, f^*(x)) + \int_X u_2^*(y) Q(\mathrm{d}y|x, f^*(x)) \text{ for all } x \in X. \tag{3.4}$$

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

(c) Any stationary policy, $f \in F$, realizing the minimum of (3.3) is average optimal; thus, f^* in (3.4) is an average optimal stationary policy.

Theorem 3.1 shows the existence of an average optimal stationary policy. Obviously, the Condition C in Theorem 3.1 is not easy to verify, whereas [6] provides some sufficient and verifiable conditions for it. We next give new assumptions for the verification of Condition C and for the existence of an average optimal stationary policy.

4. NEW CONDITIONS AND MAIN RESULTS

As is known, [6] provides quite general assumptions (i. e. Conditions A, B, and C above) to guarantee the existence of an average optimal stationary policy. The aim of our paper is to give new sufficient conditions for the verification of Conditions A–C. Since the Condition B is imposed on the primitive data (X, A, K, Q, c) of the model, our efforts are devoted to the verification of Conditions A and C, and we give new assumptions below.

Assumption 1. There exist a measurable function $\omega \ge 1$ on X, positive constants $0 < \lambda < 1,\ b_1 < \infty$, and some state $z \in X$, such that

$$\int_X Q(\mathrm{d}y|x,a)\omega(y) \le \lambda\omega(x) + b_1 \mathbb{I}_{\{z\}}(x) \quad \text{for each } (x,a) \in K, \tag{4.1}$$

where $\mathbb{I}_D$ stands for the indicator function of any set D.

Assumption 2. $|c(x,a)| \le M_1\omega(x)$ for all $(x,a) \in K$, with some constant $M_1 > 0$.

The hypothesis of (4.1) is called a Lyapunov-type condition. Note that Assumptions 1 and 2 are imposed on the primitive data (X, A, K, Q, c).

We are ready to give our main result of this paper.

Theorem 4.1. Suppose that Assumptions 1, 2, and Condition B are satisfied for the primitive data (X, A, K, Q, c). Then Conditions A and C also hold. Thus (by Theorem 3.1), there exists an average optimal stationary policy for the control model.

Proof. See below. □

Remark 4.1. (a) A key feature of Assumptions 1 and 2 is that they mainly involve the Lyapunov-type condition. However, in order to verify Condition C in [6, Lemma 3.3(b)], besides Assumptions 1, 2, the hypothesis of stochastic monotonicity of the resulting Markov chains $\{x_n\}$ is also required (for each $f \in F$).

(b) The inequality (4.1) is slightly stronger than the Lyapunov-type condition (iv) in Proposition 10.2.5 in [8] (i.e. the Condition P_4 in Remark 5.2 below). However, some additional ones such as the minorant condition (i) in Proposition 10.2.5 in [8] (i.e. the Condition P_1 in Remark 5.2 below) are needed.

(c) Since $\omega(x)$ may be unbounded on X, we see from Assumption 2 that the cost function is allowed to be unbounded from above and from below. Thus, our Theorem 4.1 is suitable for some certain models where the cost function is not necessarily nonnegative (see Example 5.1 below).

(d) We will give an example (i.e. Example 5.2 below), for which our assumptions are all satisfied, whereas some of conditions in [3, p. 187], [4], [6, Lemma 3.3], [8, p. 126] fail to hold.

To prove Theorem 4.1, we need some auxiliary facts given in the following lemmas. The first lemma provides some estimates of the expected first passage time and expected total cost during the first passage time from any initial state x to the state z. It's well worth noting that the proof of Lemma 4.2 follows the way in [11, pp. 264–266].

For every $x \in X$, let $\tau_z := \inf\{n \geq 1 : x_n = z\}$ be the first passage time of first hitting state z of the process $\{x_n\}$.

Lemma 4.2. Under Assumptions 1 and 2, the following assertions hold.

(a) $E_x^f\left[\sum_{i=0}^{\tau_z-1}\omega(x_i)\right] \leq \frac{\omega(x)+b_1}{1-\lambda}$ and $E_x^f\left[\tau_z\right] \leq \frac{\omega(x)+b_1}{1-\lambda}$ for each $x \in X$ and $f \in F$.

(b) $E_x^f\left[\sum_{i=0}^{\tau_z-1}c(x_i, f(x_i))\right] \leq M_1\frac{\omega(x)+b_1}{1-\lambda}$ for each $x \in X$ and $f \in F$.

Proof. For ease of notation, let $g(x) := (1-\lambda)\omega(x)$ for $x \in X$. Given any $x \in X$ and $f \in F$, from (4.1) we get

$$\int_X Q_f(\mathrm{d}y|x)\omega(y) \leq \omega(x) - (g(x) \wedge N) + b_1\mathbb{I}_{\{z\}}(x) \; for \; each \; N \geq 1, \tag{4.2}$$

where the symbol "$\wedge$" means taking minimum.

Let $\mathscr{F}_i := \sigma(x_0, x_1, \ldots, x_i)$ be the smallest σ-algebra generated by the first $(i+1)$ random variables $x_0, x_1, \ldots, x_i$. By (2.1) we have $\int_X Q_f(\mathrm{d}y|x_i)\omega(y) = E_x^f\left[\omega(x_{i+1})|\mathscr{F}_i\right]$. So that with the operator E_x^f, (4.2) can be rewritten as

$$E_x^f\left[\omega(x_{i+1})|\mathscr{F}_i\right] \leq \omega(x_i) - (g(x_i) \wedge N) + b_1\mathbb{I}_{\{z\}}(x_i), \quad i = 0, 1, 2, \ldots. \tag{4.3}$$

For each $n \geq 1$, define a sequence of stopping times $\tau_z^n := \min\{n, \tau_z, \inf\{k \geq 0 : \omega(x_k) \geq n\}\}$. Then clearly $\tau_z^n \uparrow \tau_z$ as $n \to \infty$, and $\sum_{i=0}^{\tau_z^n - 1} \omega(x_i) \leq n^2$. Since for any given sample path taken from Ω, τ_z^n is deterministic, we see that

$$\omega(x_{\tau_z^n}) = \omega(x_0) + \sum_{i=1}^{\tau_z^n}(\omega(x_i) - \omega(x_{i-1})) = \omega(x_0) + \sum_{i=1}^{n} \mathbb{I}_{\{\tau_z^n \geq i\}}(\omega(x_i) - \omega(x_{i-1})). \quad (4.4)$$

Noting that $\{\tau_z^n \geq i\} \in \mathscr{F}_{i-1}$, we obtain from (4.4)

$$\begin{aligned} 0 \leq E_x^f \left[\omega(x_{\tau_z^n})\right] &= E_x^f \left[\omega(x_0)\right] + E_x^f \left[\sum_{i=1}^{n} \mathbb{I}_{\{\tau_z^n \geq i\}}(\omega(x_i) - \omega(x_{i-1}))\right] \\ &= \omega(x) + \sum_{i=1}^{n} E_x^f \left[E_x^f \left[\mathbb{I}_{\{\tau_z^n \geq i\}}(\omega(x_i) - \omega(x_{i-1}))|\mathscr{F}_{i-1}\right]\right] \\ &= \omega(x) + \sum_{i=1}^{n} E_x^f \left[\mathbb{I}_{\{\tau_z^n \geq i\}}(E_x^f \left[\omega(x_i)|\mathscr{F}_{i-1}\right] - \omega(x_{i-1}))\right] \\ &\leq \omega(x) + \sum_{i=1}^{n} E_x^f \left[\mathbb{I}_{\{\tau_z^n \geq i\}}(-(g(x_{i-1}) \wedge N) + b_1 \mathbb{I}_{\{z\}}(x_{i-1}))\right] \\ &= \omega(x) + E_x^f \left[\sum_{i=1}^{\tau_z^n}(-(g(x_{i-1}) \wedge N) + b_1 \mathbb{I}_{\{z\}}(x_{i-1}))\right], \qquad (4.5) \end{aligned}$$

where the last second inequality uses (4.3).

Noting that $\sum_{i=1}^{\tau_z^n}(g(x_{i-1}) \wedge N) \leq nN$, from (4.5) we have

$$E_x^f \left[\sum_{i=1}^{\tau_z^n}(g(x_{i-1}) \wedge N)\right] \leq \omega(x) + \ b_1 E_x^f \left[\sum_{i=1}^{\tau_z^n} \mathbb{I}_{\{z\}}(x_{i-1})\right] \leq \omega(x) + b_1. \qquad (4.6)$$

Letting $n \to \infty$ and then $N \to \infty$ in (4.6), by the monotone convergence theorem and the definition of $g(x)$, the following holds

$$E_x^f \left[\sum_{i=0}^{\tau_z - 1} \omega(x_i)\right] \leq \frac{\omega(x) + b_1}{1 - \lambda},$$

which, together with Assumption 2 and the hypothesis that $\omega \geq 1$ on X, completes the proofs. □

The following lemma about the discounted optimal criterion is needed in proving Theorem 4.1.

Lemma 4.3. Under Assumptions 1 and 2, we have

$$|(1 - \alpha)V_\alpha(x, \pi)| \leq M_1\omega(x) + M_1 b_1/(1 - \lambda) \quad \text{for all } x \in X,\ \alpha \in (0, 1) \text{ and } \pi \in \Pi.$$

Proof. Fix any initial state $x \in X$ and any policy $\pi \in \Pi$. First we assert that

$$E_x^\pi[\omega(x_n)] \leq \lambda^n \omega(x) + b_1(1-\lambda^n)/(1-\lambda). \tag{4.7}$$

Indeed, for any $n \geq 1$, by (2.1) and (4.1), we obtain

$$E_x^\pi\left[\omega(x_n)|x_0, a_0, x_1, a_1, \ldots, x_{n-1}, a_{n-1}\right] = \int_X \omega(y) Q(\mathrm{d}y|x_{n-1}, a_{n-1}) \leq \lambda\omega(x_{n-1}) + b_1.$$

Hence, by taking the expectation operator E_x^π, we have

$$\begin{aligned} E_x^\pi\left[\omega(x_n)\right] &\leq \lambda E_x^\pi\left[\omega(x_{n-1})\right] + b_1 \\ &\leq \lambda^2 E_x^\pi\left[\omega(x_{n-2})\right] + b_1 + b_1\lambda \\ &\leq \cdots\cdots\cdots \\ &\leq \lambda^n \omega(x) + b_1 + b_1\lambda + \cdots + b_1\lambda^{n-1} \\ &= \lambda^n\omega(x) + b_1(1-\lambda^n)/(1-\lambda). \end{aligned}$$

So that

$$\begin{aligned} |V_\alpha(x,\pi)| &= \left| E_x^\pi\left[\sum_{n=0}^\infty \alpha^n c(x_n, a_n)\right]\right| \\ &\leq M_1 \sum_{n=0}^\infty \alpha^n E_x^\pi\left[\omega(x_n)\right] \\ &\leq M_1 \sum_{n=0}^\infty \alpha^n\left(\lambda^n\omega(x) + b_1(1-\lambda^n)/(1-\lambda)\right) \\ &\leq M_1\omega(x)/(1-\alpha) + M_1 b_1/((1-\alpha)(1-\lambda)), \end{aligned}$$

where we have used (4.7) and the hypotheses that $|c(x,a)| \leq M_1\omega(x)$ and $0 < \lambda < 1$. □

Proof of Theorem 4.1.

Since Assumptions 1 and 2 imply Condition A, we only have to show Condition C. Under Assumptions 1, 2 and Condition B, by Theorem 8.3.6 in [7] or Lemma 3.2 in [6], for each $0 < \alpha < 1$, there exists a stationary policy f_α (depending on α) satisfying $V_\alpha^*(x) = V_\alpha(x, f_\alpha)$ for all $x \in X$. For ease of notation, let $c_f(x) := c(x, f(x))$ for all $x \in X$ and $f \in F$. Using the strong Markov property, we have

$$\begin{aligned} h_\alpha(x) &= V_\alpha^*(x) - V_\alpha^*(z) \\ &= V_\alpha(x, f_\alpha) - V_\alpha(z, f_\alpha) \\ &= E_x^{f_\alpha}\left[\sum_{n=0}^{\tau_z - 1}\alpha^n c_{f_\alpha}(x_n)\right] - V_\alpha(z, f_\alpha) + E_x^{f_\alpha}\left[E_x^{f_\alpha}\left[\sum_{n=\tau_z}^\infty \alpha^n c_{f_\alpha}(x_n)|\mathcal{F}_{\tau_z}\right]\right] \\ &= E_x^{f_\alpha}\left[\sum_{n=0}^{\tau_z - 1}\alpha^n c_{f_\alpha}(x_n)\right] + V_\alpha(z, f_\alpha) E_x^{f_\alpha}\left[\alpha^{\tau_z} - 1\right]. \end{aligned} \tag{4.8}$$

Since $1-\alpha^n \leq (1-\alpha)n$ for all $n \geq 1$, by (4.8) and Lemmas 4.2 and 4.3, we have

$$\begin{aligned}
|h_\alpha(x)| &\leq E_x^{f_\alpha}\left[\sum_{n=0}^{\tau_z-1}\alpha^n|c_{f_\alpha}(x_n)|\right] + |V_\alpha(z,f_\alpha)|\, E_x^{f_\alpha}\left[|\alpha^{\tau_z}-1|\right] \\
&\leq E_x^{f_\alpha}\left[\sum_{n=0}^{\tau_z-1}\alpha^n|c_{f_\alpha}(x_n)|\right] + |(1-\alpha)V_\alpha(z,f_\alpha)|E_x^{f_\alpha}[\tau_z] \\
&\leq M_1(\omega(x)+b_1)/(1-\lambda) + M_1\left(\omega(z)+b_1/(1-\lambda)\right)(\omega(x)+b_1)/(1-\lambda) \\
&\leq \frac{M_1(1+b_1)}{1-\lambda}\left(1+\omega(z)+\frac{b_1}{1-\lambda}\right)\omega(x), \qquad (4.9)
\end{aligned}$$

where the last inequality uses the hypothesis that $\omega(x) \geq 1$. Taking

$$\begin{aligned}
v_1(x) &:= -\frac{M_1(1+b_1)}{1-\lambda}\left(1+\omega(z)+\frac{b_1}{1-\lambda}\right)\omega(x), \\
v_2(x) &:= \frac{M_1(1+b_1)}{1-\lambda}\left(1+\omega(z)+\frac{b_1}{1-\lambda}\right)\omega(x),
\end{aligned}$$

obviously we have $v_1(x)$, $v_2(x) \in \mathcal{B}_\omega(X)$. Therefore Condition C follows from (4.9). □

5. EXAMPLES

In this section, we will illustrate our results with two examples. One of the examples is on applications of our results to an admission control problem of queueing systems, and the other one is to show the difference between our conditions and those in the previous literature. As is well known, admission control problems of queueing systems have been widely studied in [12] and [14]. Here we are interested in the *new* average optimality condition for the admission control problem of a queueing system (i. e. Example 2.2.1 in [14, pp. 16–18,145–148]).

Example 5.1. (The admission control problem in [14].) Consider a controlled queueing system with the service policy being First Come First Served and the state equation:

$$X_{n+1} = X_n + \mathbb{I}_{\{incoming\ customers\ acceptted\}}Y_n - Z_n, \quad n = 0, 1, \ldots,$$

where X_n and Y_n represent the number of waiting customers in the queue and the number of incoming customers who ask for admissions into the queue at the beginning of the nth period for all $n \geq 0$ respectively, Z_n denotes the number of service completions during the nth period. When the batch of Y_n customers arrive, the decision maker has to choose one of two actions: to reject all of them, or to accept them into the queue. We assume that the Y_ns are i.i.d with the common distribution $P(Y_0{=}y){=}p_y$ for all integers $y \geq 0$. Moreover, suppose that the values of Z_n takes either 0 or 1, that a service occurs during period n according to a geometric distribution with a fixed rate $\mu \in (0,1)$ when there is at least one customer waiting in the queue (i. e. $X_n > 0$), and that no service will be provided if $X_n = 0$).

We now formulate this model as an MDP. The state space is $X := \{0, 1, 2, \ldots\}$, and the set of available actions given $x \in X$ is $A(x)$:={0,1}, where $a = 0$ standards " to reject the incoming customers" and $a = 1$ standards " to accept the incoming customers". According to the description of the model, we see that the transition kernel can be given as follows:

$$\begin{aligned} Q(0|0,0) &= 1, & Q(y|0,1) &= p_y, & y &\geq 0; \quad (5.1)\\ Q(x-1|x,0) &= \mu, & Q(x|x,0) &= 1-\mu, & x &\geq 1; \quad (5.2)\\ Q(x-1|x,1) &= \mu p_0, & Q(x+y|x,1) &= \mu p_{y+1} + (1-\mu)p_y, & x &\geq 1,\ y \geq 0. \quad (5.3) \end{aligned}$$

We denote the cost function for this example by $c(x, a)$ for all $x \in X$ and $a \in A(x)$.

We aim to find conditions that ensure the existence of an average optimal stationary policy for the admission control problem. To do so, we consider the following conditions.

$\mathbf{D}_1$: *There exists some* $\gamma > 0$ *such that* $E[e^{\gamma Y_0}] < \infty$ *and* $E[Y_0] < \mu$.

$\mathbf{D}_2$: $|c(x, a)| \leq L'(1 + x^n)$ *for all* $x \in X$ *and* $a \in A(x)$, *for some positive integer* n *and a constant* $L' > 0$.

We next state and prove the main result about the admission control problem above.

Proposition 5.1. Suppose that the Conditions $\mathbf{D}_1$, $\mathbf{D}_2$ hold. Then Assumptions 1, 2 and Condition B are satisfied for the admission control problem. Hence (by Theorem 4.1), there exists an average optimal stationary policy for Example 5.1.

Proof. To verify the inequality (4.1) in Assumption 1 for Example 5.1, let

$$\lambda := (1 - \mu + \mu e^{-\delta_0}) E\left[e^{\delta_0 Y_0}\right], \tag{5.4}$$

where $\delta_0 \in (0, \gamma]$ is a constant satisfying $0 < \lambda < 1$. To prove that such a δ_0 exists, we define $f(t) := (1 - \mu + \mu e^{-t}) E\left[e^{tY_0}\right],\ t \in [0, \gamma]$. As a result of the Condition $\mathbf{D}_1$ and Theorem 2.3.7 in [2] , the first moment of Y_0, $E[Y_0]$, equals the value of the first derivative of the moment generating function, $E\left[e^{tY_0}\right]$, at $t = 0$, i. e. $E[Y_0] = \{E[e^{tY_0}]\}'\Big|_{t=0}$. Then $f^{'}(0) = (1 - \mu + \mu e^{-t})|_{t=0} \cdot \{E[e^{tY_0}]\}'\Big|_{t=0} - \mu e^{-t}|_{t=0} \cdot E\left[e^{tY_0}\right]\big|_{t=0} = E[Y_0] - \mu$. Since $E[Y_0] < \mu$ implied by Condition $\mathbf{D}_1$, we get $f^{'}(0) < 0$. Thus we can choose a number $\delta_0 \in (0, \gamma]$ such that $f(\delta_0) < f(0) = 1$.

We aim to verify (4.1) with λ as in (5.4), $\omega(x) := e^{\delta_0 x}$, $z := 0$, and $b_1 := E\left[e^{\delta_0 Y_0}\right]$. So we can rewrite (4.1) as the following

$$\sum_{y \in X} Q(y|x, a) e^{\delta_0 y} \leq \lambda e^{\delta_0 x} + E\left[e^{\delta_0 Y_0}\right] \mathbb{I}_{\{0\}}(x) \quad \textit{for all } x \in X,\ a \in \{0, 1\}. \tag{5.5}$$

We proceed to verify (5.5) in three steps.

Case 1. If $x = 0$, by (5.1) we have that $\sum_{y \in X} Q(y|0,0)e^{\delta_0 y} = 1$ and $\sum_{y \in X} Q(y|0,1)e^{\delta_0 y} = \sum_{y=0}^{\infty} p_y e^{\delta_0 y} = E\left[e^{\delta_0 Y_0}\right]$. So (5.5) holds for $x = 0$, $a \in \{0, 1\}$.

Case 2. If $x > 0$ and the action 0 is taken, by (5.2), we have $\sum_{y\in X} Q(y|x,0)e^{\delta_0 y} = \mu e^{\delta_0(x-1)} + (1-\mu)e^{\delta_0 x} = (1-\mu+\mu e^{-\delta_0})e^{\delta_0 x} \le \lambda\omega(x)$. So (5.5) holds for $x > 0$ and $a = 0$.

Case 3. If $x > 0$ and the action 1 is taken, by (5.3), we have

$$\begin{aligned}\sum_{y\in X} Q(y|x,1)e^{\delta_0 y} &= \mu p_0 e^{\delta_0(x-1)} + \sum_{y=0}^{\infty}(\mu p_{y+1} + (1-\mu)p_y)e^{\delta_0(x+y)} \\ &= \left(\mu p_0 e^{-\delta_0} + \sum_{y=0}^{\infty}(\mu p_{y+1} + (1-\mu)p_y)e^{\delta_0 y}\right) e^{\delta_0 x} \\ &= \left(1-\mu+\mu e^{-\delta_0}\right) E\left[e^{\delta_0 Y_0}\right] e^{\delta_0 x} \\ &= \lambda\omega(x).\end{aligned}$$

From Cases 1–3, we obtain that (5.5) (i. e. (4.1)) holds for all $x \in X$ and $a \in \{0,1\}$. Thus Assumption 1 is satisfied.

Moreover, for the n as in Condition $\mathbf{D}_2$ and the above $\delta_0 > 0$, there exists a constant $l > 0$ such that $1 + x^n \le le^{\delta_0 x} = l\omega(x)$ for all $x \ge 0$. Hence by Condition $\mathbf{D}_2$, we see Assumption 2 holds.

On the other hand, since $A(x) = \{0,1\}$ only has two elements for each $x \in X$, Condition B holds naturally. Thus (by Theorem 4.1), Proposition 5.1 follows. □

As has been noted, our Example 5.1 is based on Example 2.2.1 in [14, pp. 16–18,145–148]. So we would like to compare our conditions with those therein.

Remark 5.1. (a) The cost structure in [14, pp. 16–18,145–148] is assumed nonnegative and defined explicitly as $c(x,1) = H(x)$ and $c(x,0) = H(x) + M$, where the positive constant M is interpreted as the rejection cost and $H(x)$ as the cost for keeping x customers waiting. $H(x)$ is also assumed to be increasing in x with $H(0) = 0$. But in our model, the cost function may have neither upper nor lower bounds.

(b) It is required in [14, pp. 16–18,145–148] that there is a positive probability of no arrivals and a positive probability of at least one customer arriving, i. e. $0 < p_0 < 1$, whereas in our model the distribution of Y_0, p, needs not satisfy this condition. Instead, we require that on one hand, the moment generating function of Y_0 exists in some interval $(0,\gamma)$, a condition satisfied trivially if there are only finitely many arrivals each time, and on the other hand, the expected arrivals in each period is less than μ, which is interpreted as the probability of a successful service.

Example 5.1 illustrates one application of our results above. The next example is to show the difference between our verifiable conditions and some existing ones in previous literature. First, we give a model of MDP and show that it satisfies Assumptions 1, 2 and Condition B. Thus the problem has an average optimal stationary policy. Then we proceed to prove that some of existing conditions in related literature such as the stochastic monotonicity and minorant condition fail to hold for the model; see Example 5.2 and Remark 5.2 below.

Example 5.2. Consider a controlled model with a state space $X := [0,\infty)$, the action space $A := (-\infty, m_{\theta,\rho}]$, where θ and ρ are two positive constants satisfying $m_{\theta,\rho} := \theta/2 - 1/\rho + 1 > 0$. Clearly we have $1 - x + \theta < 0$ for $x \geq \theta + 2$. The set of available actions given state x, is defined by

$$A(x) := \begin{cases} [1 - x + \theta, m_{\theta,\rho}], & x \geq \ \theta + 2, \\ \{0\}, & x \in [0, \theta + 2), \end{cases}$$

and the transition kernel is defined by the following

$$Q(D|x,a) := \begin{cases} \frac{\rho}{\theta}\mu_L(D \cap [x + a - \theta, x + a]) + (1 - \rho)\mathbb{I}_D(0), & x \geq \ \theta + 2, \\ \mathbb{I}_D(0), & x \in (0, \theta + 2), \\ \mathbb{I}_D(\theta + 2), & x = 0, \end{cases}$$

where D is any set in $\mathcal{B}(X)$, and μ_L denotes the Lebesgue measure on the real line. Obviously, the set of all state-action pairs, K, is a Borel subset of $X \times A$.

Denote the cost function of this control model by $c(x,a)$.

For the existence of an average optimal stationary policy, consider the following assumptions.

$\mathbf{F}_1$: $1/2 \leq \rho < 1$, *and* $\theta > 2$*;*

$\mathbf{F}_2$: *For each* $x \in X$*,* $c(x,a)$ *is lower-semi continuous in* $a \in A(x)$*, and* $|c(x,a)| \leq L \cdot (x + 1)$ *for all* $a \in A(x)$*, with some constant* $L > 0$*.*

Note that the Condition $\mathbf{F}_1$ ensures that $m_{\theta,\rho} > 0$.

Under the Condition $\mathbf{F}_1$, we show that there exists a function $\omega(x) \geq 1$ on X such that

$$Q\omega(x,a) := \int_X Q(y|x,a)\omega(y) \leq \rho\omega(x) + (\theta + 4)\mathbb{I}_{\{0\}}(x) \text{ for each } x \in X \textit{ and } a \in A(x), \tag{5.6}$$

which implies that Assumption 1 holds with $\lambda := \rho$, $z := 0$, $b_1 := \theta + 4$. Indeed, define $\omega(x) := x + 2$, when $x > 0$, and $\omega(0) := 1$.

When $x = 0$, (5.6) clearly holds.

When $0 < x < \theta + 2$, $Q\omega(x,a) = \omega(0) = 1$. By Condition $\mathbf{F}_1$, we have $\rho \geq 1/2$, which gives $\rho\omega(x) > 2\rho \geq 1$ together with the definition of $\omega(x)$. So that (5.6) holds for $0 < x < \theta + 2$.

When $x \geq \theta + 2$,

$$\begin{aligned} Q\omega(x,a) &= \frac{\rho}{\theta}\int_{x+a-\theta}^{x+a} \omega(y)\mu_L(\mathrm{d}y) + (1 - \rho)\omega(0) \\ &= 1 - \rho + 2\rho + \frac{\rho}{2\theta}[(x + a)^2 - (x + a - \theta)^2] \\ &= \rho x + \rho a - \rho\theta/2 + 1 + \rho, \end{aligned} \tag{5.7}$$

which, together with the hypothesis that $a \leq m_{\theta,\rho} = \theta/2 - 1/\rho + 1$, gives

$$Q\omega(x,a) \leq \rho x + \theta\rho/2 - 1 + \rho - \rho\theta/2 + 1 + \rho = \rho(x + 2) = \rho\omega(x).$$

Thus the proof of (5.6) is completed.

By Condition $\mathbf{F}_2$, we know that Assumption 2 holds since $|c(x,a)| \le L(x+1) \le L\omega(x)$ for all state-action pairs. Also, from the calculations in (5.7), we see that for every bounded measurable function u on X, $Qu(x,a)$ is continuous in $a \in A(x)$ given $x \in X$ if and only if the integral $\int_{x+a-\theta}^{x+a} u(y)\mu_L(\mathrm{d}y)$ is continuous in a. And the latter fact follows from the absolute continuity of the Lebesgue integral. The rest of the hypotheses in Condition B are easy to verify. All of the above discussions lead to the following proposition.

Proposition 5.2. Under the Conditions $\mathbf{F}_1$ and $\mathbf{F}_2$, there exists an average optimal stationary policy for Example 5.2.

To state the difference between our verifiable conditions and some existing ones in previous literature, we need to recall some concepts below.

Stochastic monotonicity (see [10], for instance): The transition kernel $Q_f(\cdot|\cdot)$ is said to satisfy the stochastic monotonicity condition, if $x < y$ implies $P_x^f(x_n \le z) \ge P_y^f(x_n \le z)$ for every $n \ge 1$ and $z \in X$, where $\{x_n\}$ is the Markov chain with the transition law $Q_f(\cdot|\cdot)$.

The verifiable uniformly ω-exponentially ergodicity conditions (in [4], [8, p. 126], or [6, Lemma 3.3(d)]: For each $f \in F$, there exist a measurable function, l_f, $0 \le l_f \le 1$ (depending on f), on X, a probability measure, ϑ, on X, and constants, $\delta_2 > 0$, $0 < \beta_2 < 1$, independent of f, such that

$\mathbf{P}_1:$ $Q(C|x,f(x)) \ge l_f(x)\vartheta(C)$ for all $C \in \mathcal{B}(X)$ and $x \in X$,

$\mathbf{P}_2:$ $\int_X l_f(y)\vartheta(\mathrm{d}y) \ge \delta_2$,

$\mathbf{P}_3:$ $\vartheta(\omega) := \int_X \omega(y)\vartheta(\mathrm{d}y) < \infty$,

$\mathbf{P}_4:$ $\int_X \omega(y)Q(\mathrm{d}y|x,f(x)) \le \beta_2\omega(x) + l_f(x)\vartheta(\omega)$ for all $x \in X$.

Remark 5.2. (a) *For any $f \in F$ the transition kernel $Q_f(\cdot|\cdot)$ in Example 5.2 is not stochastic monotone, therefore, the conditions in [6, Lemma 3.3(b)] are not satisfied.*

Indeed, by the construction of the transition function, we have $P_0(x_1 \le 1/2) = 0$, and $P_1(x_1 \le 1/2) = 1$, so that $P_0(x_1 \le 1/2) < P_1(x_1 \le 1/2)$, which contradicts with the definition of stochastic monotonicity.

(b) *The earlier conditions for the uniformly ω-exponentially ergodicity in [8, p. 126] or [6, Lemma 3.3(d)] fail to hold.*

We next prove the assertion in Remark 5.2 (b) by contradiction as follows. Let $f^*(x)$: $X \to A$ denotes such a decision rule: $f^*(x) = 1 - x + \theta$, when $x \ge \theta + 2$; $f^*(x) = 0$, when $x < \theta + 2$. Suppose that $l_{f^*}(x)$ satisfies the Conditions $\mathbf{P}_1$–$\mathbf{P}_4$ above. Then, the following facts hold:

(1) If $l_{f^*}(0) > 0$, then, since $Q(\{\theta+2\}|0,0) = 1$, we get $\vartheta(\{\theta+2\}^c) = 0$. For if not so, replace C in Condition $\mathbf{P}_1$ with $\{\theta+2\}^c$, and it follows that $Q(\{\theta+2\}^c|0,0) \ge l_{f^*}(0)\vartheta(\{\theta+2\}^c) > 0$ (contradictory).

From Condition $\mathbf{P}_2$, $l_{f^*}(\theta+2) \geq \delta_2 > 0$. Replacing C in Condition $\mathbf{P}_1$ with $\{\theta+2\}$, and x with $(\theta+2)$, we have $Q(\{\theta+2\}|\theta+2, f^*(\theta+2)) \geq l_{f^*}(\theta+2)\vartheta(\{\theta+2\}) > 0$, which contradicts with the fact that $Q(\{\theta+2\}|\theta+2, a) = 0$ for all $a \in A(\theta+2)$. Therefore, $l_{f^*}(0) = 0$.

(2) If $l_{f^*}(\theta+2) > 0$, we know from the construction of Q that $Q(\{0\} \cup [1, \theta+1]|\theta+2, f^*(\theta+2)) = 1$. As the arguments in part (1), we can prove that $\vartheta(\{\{0\} \cup [1, \theta+1]\}^c) = 0$. By Condition $\mathbf{P}_2$ and $l_{f^*}(0) = 0$, the following hold: $\vartheta([1, \theta+1]) > 0$ and there exists at least an element $x_0 \in [1, \theta+1]$ such that $l_{f^*}(x_0) > 0$. By Condition $\mathbf{P}_1$, we have $Q([1, \theta+1]|x_0, f^*(x_0)) \geq l_{f^*}(x_0)\vartheta([1, \theta+1]) > 0$ (contradictory). Therefore, $l_{f^*}(\theta+2) = 0$.

(3) If there exists one number (denoted by y) in $(0, \theta+2)$ s.t. $l_{f^*}(y) > 0$, following the same route as in the previous proof, we get $\vartheta(\{0\}^c) = 0$. Using $l_{f^*}(0) = 0$, we have $\int_X l_{f^*}(x)v(\mathrm{d}x) = l_{f^*}(0)v(\{0\}) = 0$, which contradicts with l_{f^*} satisfying Condition $\mathbf{P}_2$.

Thus, it follows from $(1)-(3)$ that $l_{f^*}(x) \equiv 0$ for all $x \in [0, \theta+2]$. This fact together with Condition $\mathbf{P}_4$ gives $\int_X \omega(y)Q(\mathrm{d}y|x, f(x)) \leq \beta_2\omega(x)$ for all $0 \leq x \leq \theta+2$. Using the definition of Q, we have

$$\omega(\theta+2) \leq \beta_2\omega(0), \tag{5.8}$$

$$\omega(0) \leq \beta_2\omega(x) \quad \forall\, x \in (0, \theta+2), \tag{5.9}$$

$$(1-\rho)\omega(0) + \frac{\rho}{\theta}\int_1^{\theta+1} \omega(y)\mu_L(\mathrm{d}y) \leq \beta_2\omega(\theta+2). \tag{5.10}$$

Since (5.9) gives $\omega(x) > \omega(0)$ for all $x \in (0, \theta+2)$, replacing $\omega(y)$ in (5.10) with $\omega(0)$, we get $\omega(0) < \omega(\theta+2)$. By (5.8), it follows that $\omega(\theta+2) < \omega(\theta+2)$ (contradictory). So, it's impossible to find a function $l_{f^*}(x)$ satisfying the Conditions $\mathbf{P}_1$–$\mathbf{P}_4$.

(c) *The minorant condition in [3, p. 187] fails to hold.*

Recall the definition of the minorant condition in [3, p. 187]: there exists a nontrivial measure v such that $Q(\Gamma|x, a) > v(\Gamma)$ for all $x \in X,\ a \in A(x),\ \Gamma \in \mathcal{B}(X)$. The proof of the assertion in Remark 5.2 (c) is similar to that in Remark 5.2 (b).

(d) *The condition in [6, Lemma 3.3(c)] fails to hold, either.*

One of assumptions in [6, Lemma 3.3(c)] assumes that there exists a positive number γ such that for each $f \in F$, there exists an atom, α_f (depending on f), such that $Q_f(\alpha_f|x) > \gamma$ for all $x \in \alpha_f$. We prove below that such an atom does not exist. According to the construction of Q, there are three kinds of set that may be atoms: a set made up of a single state, a subset of $(0, \theta+2)$, and a subset of $[\theta+2, \infty)$.

Firstly, α_f cannot be taken as any single state in X or any subset of $(0, \theta+2)$, since $Q_f(\{x\}|x) = 0$ for each $x \in X$, $f \in F$ and $Q_f(\{0\}|x) = 1$ for each $x \in (0, \theta+2)$, $f \in F$.

To prove that α_f cannot be a subset of $[\theta+2, \infty)$, we consider the particular case when $f = f^*$ defined in part (b). Then from the definition of the transition kernel we have $Q_{f^*}(\{0\} \cup [1, \theta+1]|x) = 1$ for every $x \in [\theta+2, \infty)$. That contradicts with the requirement that $Q_{f^*}(\alpha_{f^*}|x) > 0$.

Remark 5.3. From Remark 5.2(a) − (d), we conclude that our conditions are indeed different from those in [3, 4, 6, 8].

(Received May 4, 2014)

REFERENCES

[1] A. Arapostathis, et al: Discrete time controlled Markov processes with average cost criterion: a survey. SIAM J. Control Optim. *31* (1993), 282–344. DOI:10.1137/0331018

[2] G. Casella and R. L. Berger: Statistical Inference. Second edition. Duxbury Thomson Learning 2002.

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

[4] E. Gordienko and O. Hernández-Lerma: Average cost Markov control processes with weighted norms: existence of canonical policies. Appl. Math. (Warsaw) *23* (1995), 2, 199–218.

[5] X. P. Guo and P. Shi: Limiting average criteria for nonstationary Markov decision processes. SIAM J. Optim. *11* (2001), 4, 1037–1053. DOI:10.1137/s1052623499355235

[6] X. P. Guo and Q. X. Zhu: Average optimality for Markov decision processes in Borel spaces: A new condition and approach. J. Appl. Probab. *43* (2006), 318–334. DOI:10.1239/jap/1152413725

[7] O. Hernández-Lerma and J. B. Lasserre: Discrete-Time Markov Control Processes. Springer, New York 1996. DOI:10.1007/978-1-4612-0729-0

[8] O. Hernández-Lerma and J. B. Lasserre: Further Topics on Discrete-Time Markov Control Processes. Springer, New York 1999. DOI:10.1007/978-1-4612-0561-6

[9] M. Kakumanu: Nondiscounted continuous time Markov decision process with countable state space. SIAM J. Control Optim. *10* (1972), 1, 210–220. DOI:10.1137/0310016

[10] R. B. Lund and R. L. Tweedie: Geometric convergence rates for stochastically ordered Markov chains. Math. Oper. Res. *21* (1996), 1, 182–194. DOI:10.1287/moor.21.1.182

[11] S. P. Meyn and R. L. Tweedie: Markov Chains and Stochastic Stability. Cambridge Univ. Press, New York 2009. DOI:10.1017/cbo9780511626630

[12] M. L. Puterman: Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley, New York 1994. DOI:10.1002/9780470316887

[13] L. I. Sennott: Average reward optimization theory for denumerable state spaces. In: Handbook of Markov Decision Processes (Int. Ser. Operat. Res. Manag. Sci. 40) (E. A. Feinberg and A. Shwartz Kluwer, eds.), Boston, pp. 153–172. DOI:10.1007/978-1-4615-0805-2_5

[14] L. I. Sennott: Stochastic Dynamic Programming and the Control of Queueing Systems. Wiley, New York 1999. DOI:10.1002/9780470317037

[15] Q. X. Zhu: Average optimality for continuous-time jump Markov decision processes with a policy iteration approach. J. Math. Anal. Appl. *339* (2008), 1, 691–704. DOI:10.1016/j.jmaa.2007.06.071

Xiaolong Zou, School of Mathematics and Computational Science, Sun Yat-sen University, Guangzhou, 510275. P. R. China.
e-mail: sysuzxl@126.com

Xianping Guo, School of Mathematics and Computational Science, Sun Yat-sen University, Guangzhou, 510275. P. R. China.
e-mail: mcsgxp@mail.sysu.edu.cn