Qingda Wei; Xian Chen

Strong average optimality criterion for continuous-time Markov decision processes

Kybernetika, Vol. 50 (2014), No. 6, 950–977

Persistent URL: http://dml.cz/dmlcz/144118

KYBERNETIKA — VOLUME *50* (2014), NUMBER 6, PAGES 950–977

STRONG AVERAGE OPTIMALITY CRITERION FOR CONTINUOUS-TIME MARKOV DECISION PROCESSES

Qingda Wei and Xian Chen

This paper deals with continuous-time Markov decision processes with the unbounded transition rates under the strong average cost criterion. The state and action spaces are Borel spaces, and the costs are allowed to be unbounded from above and from below. Under mild conditions, we first prove that the finite-horizon optimal value function is a solution to the optimality equation for the case of uncountable state spaces and unbounded transition rates, and that there exists an optimal deterministic Markov policy. Then, using the two average optimality inequalities, we show that the set of all strong average optimal policies coincides with the set of all average optimal policies, and thus obtain the existence of strong average optimal policies. Furthermore, employing the technique of the skeleton chains of controlled continuous-time Markov chains and Chapman–Kolmogorov equation, we give a new set of sufficient conditions imposed on the primitive data of the model for the verification of the uniform exponential ergodicity of continuous-time Markov chains governed by stationary policies. Finally, we illustrate our main results with an example.

Keywords: continuous-time Markov decision processes, strong average optimality criterion, finite-horizon expected total cost criterion, unbounded transition rates, optimal policy, optimal value function

Classification: 93E20, 90C40

1. INTRODUCTION

Continuous-time Markov decision processes (CTMDPs) have been deeply studied under different optimality criteria in recent years. As is well known, the expected average criterion is one of the most common optimality criteria, and the existence of average optimal policies for CTMDPs has been studied via different methods and sets of conditions; see, for instance, [11, 13, 14, 21, 23, 24] and the references therein. However, the expected average criterion is rather underselective due to the fact that it neglects the behavior of the controlled stochastic process during any finite time interval. Therefore, some advanced optimality criteria, such as the bias, weakly overtaking and variance minimization criteria, have been proposed; see [13] for details. Motivated by the strong average optimality criterion for discrete-time MDPs in [3, 7, 8, 15], which evaluates the performance of a policy over long but finite horizons, as well as in the long-run average

DOI: 10.14736/kyb-2014-6-0950

sense, we are concerned with the continuous-time version (see Definition 2.2) in this paper. To the best of our knowledge, there is no literature dealing with the strong average optimality criterion for CTMDPs. As indicated in [3, 7, 8, 15], every strong average optimal policy is average optimal under the nonnegativity assumption on the costs, but the contrary is not necessarily true without further conditions. Consequently, it is desirable for us to study the relation between the average optimality and strong average optimality. Moreover, the strong average optimality criterion provides a *new* way to overcome the underselective deficiency of the expected average criterion. It should be mentioned that we discuss the strong average optimality criterion in the class of all randomized Markov policies whereas the advanced optimality criteria studied in [13] are restricted to the class of all deterministic stationary policies.

In this paper, we study the strong average criterion for CTMDPs with the *unbounded* transition rates in which the state and action spaces are Borel spaces, and the costs are allowed to be *unbounded from above and from below*. Since the definition of the strong average criterion involves the finite-horizon expected total cost criterion, we also need to investigate the existence of optimal policies for CTMDPs under the finite-horizon criterion, whose treatment is more complicated than that for discrete-time MDPs. The finite-horizon criterion for CTMDPs has been studied by many authors; see, for instance, [1, 4, 18] for the case of finite or denumerable states, and [9, 10, 19, 22] for the case of a Borel state space. As can be seen in the previous literature, the common approach to study the finite-horizon criterion for CTMDPs is via establishing the optimality equation, and they all deal with the case of bounded transition rates except [4]. It should be noted that the uniformization method is inapplicable in this paper because the transition rates are allowed to be *unbounded*. Under mild conditions, following the technique of time-discretization used in [4], we extend the optimality equation for finite-horizon criterion to the case of *uncountable* state spaces and *unbounded* transition rates. Then, we show that the finite-horizon optimal value function is a solution to the optimality equation, and that there exists an optimal deterministic Markov policy, which have not been proven in [4] (see Theorem 4.1 and Remark 4.2).

Basing on the two average optimality inequalities established in [11] and the existence of optimal policies for finite-horizon criterion, under suitable conditions, we show that the set of all strong average optimal policies coincides with the set of all average optimal policies by using the Kolmogorov forward equation, and thus obtain the existence of a strong average optimal stationary policy (see Theorem 4.3 and Remark 4.4). Furthermore, as we can see from the existing works on the expected average criterion for CTMDPs, the assumption that the relative difference of the discount optimal value function is bounded by an integrable function (see Assumption 3.2), which is weaker than the uniform exponential ergodicity condition in [23, 24], plays a crucial role in ensuring the existence of average optimal policies. However, it is difficult to verify this assumption because it does not impose on the primitive data of the model. Thus, it is necessary to give some sufficient conditions for the verification of this assumption; see the discussions in [11, 21]. In this paper, we give a *new* set of *verifiable* sufficient conditions imposed on the *primitive data* of the model for the verification of the uniform ω-exponential ergodicity of continuous-time Markov chains governed by stationary policies (see Theorem 4.5 and Remark 4.6). More precisely, inspired by Theorem 2.3 in [17] concerning the

uniform ω-geometrical ergodicity of discrete-time Markov chains, we obtain the uniform ω-geometrical ergodicity of some skeleton chains of controlled continuous-time Markov chains by employing the construction of the transition function with the corresponding transition rates. Then, from the geometrical ergodicity of the skeleton chains and Chapman–Kolmogorov equation, we show that our *new* set of sufficient conditions implies the uniform ω-exponential ergodicity of controlled continuous-time Markov chains.

The rest of this paper is organized as follows. In Section 2, we introduce the control model and optimality criteria. In Section 3, we give optimality conditions for the existence of optimal policies and some preliminary lemmas. In Section 4, we state and prove our main results. In Section 5, we illustrate our main results with an example.

2. THE MODEL AND OPTIMALITY CRITERIA

The control model of CTMDPs under consideration is as follows:

$$\{X, A, (A(x), x \in X), q(\cdot|x,a), c(x,a)\},$$

where X and A are state and action spaces, which are assumed to be Borel spaces with Borel σ-algebras $\mathcal{B}(X)$ and $\mathcal{B}(A)$, respectively. $A(x) \in \mathcal{B}(A)$ denotes the set of admissible actions at the state $x \in X$. Let $K := \{(x,a)|\ x \in X, a \in A(x)\}$, and assume that K is a measurable subset of $X \times A$ and contains the graph of a measurable mapping from X to A. The transition rates $q(\cdot|x,a)$ are supposed to satisfy the following properties:

- For each fixed $(x,a) \in K$, $q(\cdot|x,a)$ is a signed measure on $\mathcal{B}(X)$, and for each fixed $D \in \mathcal{B}(X)$, $q(D|\cdot)$ is a real-valued Borel-measurable function on K;
- $0 \le q(D|x,a) < \infty$ for all $(x,a) \in K$ and $x \notin D \in \mathcal{B}(X)$;
- $q(X|x,a) = 0$ for all $(x,a) \in K$;
- $q^*(x) := \sup_{a \in A(x)} |q(\{x\}|x,a)| < \infty$ for all $x \in X$.

Finally, $c(x,a)$, a real-valued cost function, is Borel-measurable on K.

To precisely define the optimality criteria, we need to introduce the concept of a policy.

Definition 2.1. A *randomized Markov policy* is a family $\pi := \{\pi_t, t \ge 0\}$ of stochastic kernels that satisfy

(i) for each $t \ge 0$, π_t is a stochastic kernel on A given X such that $\pi_t(A(x)|x) = 1$ for all $x \in X$;

(ii) for each $D \in \mathcal{B}(A)$, $\pi_t(D|x)$ is Borel-measurable in $(t,x) \in [0,\infty) \times X$.

A policy π is said to be *(deterministic) Markov* if there exists a Borel-measurable function f on $[0,\infty) \times X$ with $f(t,x) \in A(x)$, such that $\pi_t(\cdot|x)$ is the Dirac measure at $f(t,x) \in A(x)$ for all $x \in X$ and $t \ge 0$. A policy π is said to be *(deterministic) stationary* if there exists a Borel-measurable function f on X with $f(x) \in A(x)$ for all $x \in X$, such that $\pi_t(\cdot|x)$ is the Dirac measure at $f(x) \in A(x)$ for all $x \in X$ and $t \ge 0$.

We denote by Π, Π_d and F the classes of all policies, deterministic Markov policies and stationary policies, respectively. Obviously, $F \subset \Pi_d \subset \Pi$.

To guarantee the regularity of the q-processes, we need the following drift condition from [11, 12, 13, 21, 23, 24].

Assumption 2.1. There exist a measurable function $\omega \geq 1$ on X, and constants $\rho_1 > 0$, $b_1 > 0$, and $L > 0$ such that

(i) $\int_X \omega(y) q(\mathrm{d}y|x,a) \leq -\rho_1 \omega(x) + b_1$ for all $(x,a) \in K$.

(ii) $q^*(x) \leq L\omega(x)$ for all $x \in X$.

Fix an initial state $x \in X$, and an initial time $s \geq 0$. Then under Assumption 2.1, for each $\pi \in \Pi$, there exist the unique probability measure $P^\pi_{s,x}$ on some measurable space $(\Omega, \mathcal{B}(\Omega))$ and a stochastic process $\{x(t), t \geq s\}$ such that

$$P^\pi_{s,x}(x(t) \in D) = p_\pi(s,x,t,D)$$

for all $D \in \mathcal{B}(X)$ and $t \geq s \geq 0$, where $p_\pi(s,x,t,\cdot)$ denotes the transition function with transition rates $q(\cdot|x,\pi_t) := \int_{A(x)} q(\cdot|x,a)\pi_t(\mathrm{d}a|x)$. The expectation operator with respect to $P^\pi_{s,x}$ is denoted by $E^\pi_{s,x}$. If $s = 0$, we write $P^\pi_{s,x}$ and $E^\pi_{s,x}$ as P^π_x and E^π_x, respectively.

Fix a discount factor $\alpha > 0$. For each $x \in X$ and $\pi \in \Pi$, we define the *expected discounted cost* $V_\alpha(x,\pi)$ and *expected average cost* $J(x,\pi)$ as

$$V_\alpha(x,\pi) := E^\pi_x\left[\int_0^\infty \int_A e^{-\alpha t} c(x(t),a)\pi_t(\mathrm{d}a|x(t))\,\mathrm{d}t\right] \text{ and}$$

$$J(x,\pi) := \limsup_{T\to\infty} \frac{1}{T} E^\pi_x\left[\int_0^T \int_A c(x(t),a)\pi_t(\mathrm{d}a|x(t))\,\mathrm{d}t\right],$$

respectively. The corresponding *discount and average optimal value* functions are defined as

$$V^*_\alpha(x) := \inf_{\pi\in\Pi} V_\alpha(x,\pi), \quad \text{and} \quad J^*(x) := \inf_{\pi\in\Pi} J(x,\pi) \quad \text{for all } x \in X,$$

respectively. Furthermore, for each $x \in X$ and $\pi \in \Pi$, the *expected total cost from time $s \geq 0$ to the terminal time $T > 0$* is defined as

$$V_{T-s}(x,\pi) := E^\pi_{s,x}\left[\int_s^T \int_A c(x(t),a)\pi_t(\mathrm{d}a|x(t))\,\mathrm{d}t\right],$$

and the corresponding *finite-horizon optimal value* function is given by

$$V^*_T(x,s) := \inf_{\pi\in\Pi} V_{T-s}(x,\pi) \quad \text{for all} \quad s \in [0,T] \quad \text{and} \quad x \in X.$$

Definition 2.2. A policy $\pi^* \in \Pi$ is said to be

- *discount optimal* if $V_\alpha(x,\pi^*) = V^*_\alpha(x)$ for all $x \in X$;

- *average optimal* if $J(x,\pi^*) = J^*(x)$ for all $x \in X$;
- *finite-horizon optimal* if $V_T(x,\pi^*) = V_T^*(x,0)$ for all $x \in X$;
- *strong average optimal* if

$$\limsup_{T\to\infty} \frac{1}{T}[V_T(x,\pi^*) - V_T^*(x,0)] = 0 \quad \text{for all } x \in X. \tag{2.1}$$

Remark 2.3. Under Assumptions 2.1 and 3.1(i) below, by Definition 2.2, we see that every strong average optimal policy is average optimal. Indeed, suppose that π^* is strong average optimal. Then, (2.1), together with the inequality $V_T(x,\pi^*) \geq V_T^*(x,0)$ for all $x \in X$, implies

$$\lim_{T\to\infty} \frac{1}{T}[V_T(x,\pi^*) - V_T^*(x,0)] = 0 \quad \text{for all } x \in X,$$

which gives

$$\limsup_{T\to\infty} \frac{1}{T}[V_T(x,\pi^*) - V_T(x,\pi)] = \limsup_{T\to\infty} \frac{1}{T}[V_T^*(x,0) - V_T(x,\pi)] \leq 0$$

for all $x \in X$ and $\pi \in \Pi$. Hence, π^* is average optimal.

There are three main goals in this paper: (i) We will show that the finite-horizon optimal value function is a solution to the optimality equation for the case of *uncountable* state spaces and *unbounded* transition rates, and the existence of optimal policies; (ii) We will give conditions for the existence of strong average optimal policies; (iii) We will present a *new* set of sufficient conditions imposed on the primitive data of the model for the verification of the uniform ω-exponential ergodicity of controlled continuous-time Markov chains.

3. PRELIMINARIES

In this section, we give optimality conditions for the existence of optimal policies and some preliminary lemmas needed to prove our main results.

Assumption 3.1. (i) There exist constants $\rho_2 > 0$, $\rho_3 > 0$, $b_2 \geq 0$, $b_3 \geq 0$, and $M > 0$ such that $|c(x,a)| \leq M\omega(x)$,

$$\int_X \omega^2(y)q(\mathrm{d}y|x,a) \leq \rho_2\omega^2(x) + b_2, \quad \text{and} \quad \int_X \omega^3(y)q(\mathrm{d}y|x,a) \leq \rho_3\omega^3(x) + b_3$$

for all $(x,a) \in K$, where ω comes from Assumption 2.1.

(ii) For each $x \in X$, the set $A(x)$ is compact.

(iii) For each fixed $x \in X$, the functions $c(x,a)$, $\int_X \omega(y)q(\mathrm{d}y|x,a)$, and $\int_X u(y)q(\mathrm{d}y|x,a)$ are continuous in $a \in A(x)$ for all bounded measurable function u on X.

Remark 3.1. Assumption 3.1 is the finiteness and standard continuity-compactness conditions, and has been widely used for CTMDPs; see, for instance, [11, 12, 13, 21, 23, 24].

To state our third hypothesis, we need to introduce the concept of the weighted norm used in [11, 12, 13, 16, 21, 23, 24]. Let $\omega \geq 1$ be as in Assumption 2.1, and define the norm $\|u\|_\omega := \sup_{x\in X} \frac{|u(x)|}{\omega(x)}$. $B_\omega(X)$ denotes the set of all real-valued measurable functions on X with finite norm.

Assumption 3.2. There exist a function $v \in B_\omega(X)$ and some state $\widehat{x} \in X$ such that

$$|h_\alpha(x)| \leq v(x) \text{ for all } x \in X \text{ and } \alpha > 0,$$

where $h_\alpha(x) := V^*_\alpha(x) - V^*_\alpha(\widehat{x})$ is the so-called relative difference of the discount optimal value function V^*_α.

Remark 3.2. Assumption 3.2 has been used in [11, 21] to ensure the existence of average optimal policies, and is weaker than the uniform ω-exponential ergodicity condition in [23, 24]. However, since this assumption does not impose on the primitive data of the model, it is difficult to verify it. Different sets of sufficient conditions for the verification of this assumption have been given in [11, 21] as well. It should be mentioned that we give a *new* set of *verifiable* sufficient conditions imposed on the primitive data of the model for the verification of it (see Theorem 4.5).

Before stating our main result on the finite-horizon expected total cost criterion, we need some preliminary lemmas. To do so, we introduce the notation below.

Choose a measurable function m on X satisfying $m \in B_\omega(X)$ and $m(x) > q^*(x)$ for all $x \in X$. For each $(x, a) \in K$, $D \in \mathcal{B}(X)$, and $h > 0$, define

$$\begin{aligned} P(D|x,a) : &= \frac{q(D|x,a)}{m(x)} + I_D(x), \\ P_h(D|x,a) : &= [hm(x) \wedge 1]P(D|x,a) + \{1 - [hm(x) \wedge 1]\}I_D(x) \\ &= [hm(x) \wedge 1]\frac{q(D|x,a)}{m(x)} + I_D(x), \end{aligned} \tag{3.1}$$

where $y_1 \wedge y_2 := \min\{y_1, y_2\}$, and $I_D(\cdot)$ denotes the indicator function of the set D. Obviously, we see that for each fixed $(x, a) \in K$ and $h > 0$, $P(\cdot|x, a)$ and $P_h(\cdot|x, a)$ are probability measures on $\mathcal{B}(X)$. Thus, for each $h > 0$, we obtain a discrete-time MDP model M_h as follows:

$$\{X, A, (A(x), x \in X), P_h(\cdot|x, a), hc(x, a)\}.$$

We denote by $\widetilde{\Pi}_d$ the class of all deterministic Markov policies for the discrete-time MDP; see [1, 15, 16, 20] for the detailed definition. Hence, for any $\pi \in \widetilde{\Pi}_d$ and any initial state $x \in X$, the well-known Tulcea theorem [15, p. 178] gives the existence of the unique probability measure $\widetilde{P}^\pi_x$ on $(X^\infty, \mathcal{B}(X^\infty))$ and there exists a stochastic process $\{x_n, n = 0, 1, \ldots\}$ associated with the model M_h. The expectation operator with respect to $\widetilde{P}^\pi_x$ is denoted by $\widetilde{E}^\pi_x$. Moreover, we denote by $\widetilde{P}^\pi_{n,x}$ the conditional probability $\widetilde{P}^\pi_{n,x}(\cdot) := \widetilde{P}^\pi(\cdot|x_n = x)$ and $\widetilde{E}_{n,x}$ is the corresponding expectation operator.

For each $z \in (-\infty, \infty)$, define $\lfloor z \rfloor := \max\{n \in \mathbb{Z} |\ n \leq z\}$, where $\mathbb{Z}$ denotes the set of all integers. For each $h > 0$, let $N := \lfloor Th^{-1} \rfloor$. For each $n = 0, 1, \ldots, N-1$, $x \in X$

and $\pi = \{f_k, k = 0, 1, \ldots\} \in \widetilde{\Pi}_d$, we define the expected total cost from time n to time $N-1$ and the corresponding optimal value function associated with the model M_h as

$$V_n^h(x,\pi) := \widetilde{E}_{n,x}^{\pi}\left[\sum_{k=n}^{N-1} hc(x_k, f_k(x_k))\right] \quad \text{and} \quad V_n^h(x) := \inf_{\pi\in\widetilde{\Pi}_d} V_n^h(x,\pi), \tag{3.2}$$

respectively. Define $V_N^h(x) := 0$ for all $x \in X$. Then under Assumptions 2.1 and 3.1, it is well known that the sequence $\{V_n^h |\ n = 0, 1, \ldots, N\}$ satisfies

$$V_n^h(x) = \inf_{a\in A(x)}\left\{hc(x,a) + \int_X V_{n+1}^h(y) P_h(\mathrm{d}y|x,a)\right\} \tag{3.3}$$

for all $x \in X$ and $n = 0, 1, \ldots, N-1$; see [1, 2, 5, 15, 20] for details. Replacing h with $h/2$ in the model M_h, we obtain the model denoted by $M_{h/2}$. For each $n = 0, 1, \ldots, \lfloor 2Th^{-1}\rfloor$, similar to (3.2), we can define the optimal value function $V_n^{(h/2)}$ on X associated with the model $M_{h/2}$ and obtain the similar result as in (3.3).

Next, following the technique used in [4], we have the three lemmas below, which extend the results in [4] for denumerable states and actions to the case of Borel spaces.

Lemma 3.3. Under Assumptions 2.1 and 3.1, we have

(a) For each $h > 0$, $\int_X \omega(y) P_h(\mathrm{d}y|x,a) \le (1 + b_1 h)\omega(x)$ for all $(x,a) \in K$.

(b) $V_n^h \in B_\omega(X)$ and $\|V_n^h\|_\omega \le MTe^{b_1 T}$ for all $h > 0$ and $n = 0, 1, \ldots, N$, with M as in Assumption 3.1.

P r o o f. (a) By Assumption 2.1 and (3.1), a straightforward calculation yields

$$\begin{aligned}\int_X \omega(y) P_h(\mathrm{d}y|x,a) &= [hm(x) \wedge 1]\left\{\frac{1}{m(x)}\int_X \omega(y) q(\mathrm{d}y|x,a)\right\} + \omega(x)\\ &\le [hm(x) \wedge 1]\frac{1}{m(x)}(-\rho_1\omega(x) + b_1) + \omega(x)\\ &\le (1 + b_1 h)\omega(x)\end{aligned}$$

for all $(x,a) \in K$, and so part (a) holds.

(b) It follows from the measurable selection theorem in [16, p. 50] that V_n^h is measurable with respect to $\mathcal{B}(X)$ for all $h > 0$ and $n = 0, 1, \ldots, N$. Moreover, we have

$$|V_{N-l}^h(x)| \le hM\sum_{k=0}^{l-1}(1 + b_1 h)^k \omega(x) \tag{3.4}$$

for all $x \in X$, $h > 0$ and $l = 1, 2, \ldots, N$. In fact, by (3.3) and Assumption 3.1, we obtain

$$|V_{N-1}^h(x)| \le h|c(x,a)| \le hM\omega(x)$$

for all $x \in X$ and $h > 0$, and so (3.4) is true for $l = 1$. Suppose that (3.4) holds for some $l \geq 1$. Then, using (3.3) again, we have

$$|V^h_{N-(l+1)}(x)| \quad \leq \quad h|c(x,a)| + \int_X |V^h_{N-l}(y)| P_h(\mathrm{d}y|x,a),$$

which, together with Assumption 3.1, part (a) and the induction hypothesis, gives

$$\begin{aligned} |V^h_{N-(l+1)}(x)| &\leq hM\omega(x) + hM \sum_{k=0}^{l-1} (1+b_1 h)^k \int_X \omega(y) P_h(\mathrm{d}y|x,a) \\ &\leq hM \sum_{k=0}^{l} (1+b_1 h)^k \omega(x) \end{aligned}$$

for all $x \in X$ and $h > 0$, and so (3.4) follows from the induction. Thus, by (3.4) and the inequality $1 + z \leq e^z$ for all $z > 0$, we have

$$|V^h_{N-l}(x)| \leq hlMe^{b_1 hl}\omega(x) \leq MTe^{b_1 T}\omega(x)$$

for all $x \in X$, $h > 0$ and $l = 1, 2, \ldots, N$. Hence, we get the desired result. □

Lemma 3.4. Under Assumptions 2.1 and 3.1, the following inequality holds:

$$\begin{aligned} V^{(h/2)}_{2n}(x) \leq & V^h_n(x) + \sum_{l=0}^{N-n-1} L^* h^2 (1 + \rho_3 h + b_3 h)^l \omega^3(x) \\ & + L^* h (1 + \rho_3 h + b_3 h)^{N-n} \omega^3(x) \end{aligned} \tag{3.5}$$

for all $x \in X$, $h > 0$ and $n = 0, 1, \ldots, N$, where the constant $L^* := M + [M + \frac{1}{4} MTe^{b_1 T}(\rho_2 + b_2 + 2L) + MTe^{b_1 T}\|m\|_\omega](b_1 + 2L)$ is independent of h and n.

P r o o f . We will show this lemma by induction. For $n = N$, we have

$$V^{(h/2)}_{2N}(x) - V^h_N(x) = I_{\{2N < \lfloor T(h/2)^{-1} \rfloor\}} \inf_{a \in A(x)} \{(h/2)c(x,a)\} \leq hM\omega(x)$$

for all $x \in X$. Thus, (3.5) is true for $n = N$. Assume that (3.5) holds for some $n = k+1$. Then, using (3.3), we obtain

$$V^{(h/2)}_{2k}(x) = \inf_{a \in A(x)} \left\{ \frac{h}{2} c(x,a) + \int_X V^{(h/2)}_{2k+1}(y) P_{\frac{h}{2}}(\mathrm{d}y|x,a) \right\},$$

which implies

$$\begin{aligned} & V^{(h/2)}_{2k}(x) \\ \leq & \inf_{a \in A(x)} \left\{ \frac{h}{2} c(x,a) + \int_X \left[\frac{h}{2} c(y,a) + \int_X V^{(h/2)}_{2k+2}(z) P_{\frac{h}{2}}(\mathrm{d}z|y,a) \right] P_{\frac{h}{2}}(\mathrm{d}y|x,a) \right\} \end{aligned}$$

$$= \inf_{a\in A(x)}\Big\{hc(x,a)+\int_X V_{2k+2}^{(h/2)}(y)P_h(\mathrm{d}y|x,a)+\frac{h}{2}\int_X[c(y,a)-c(x,a)]P_{\frac{h}{2}}(\mathrm{d}y|x,a)$$
$$+\int_X\int_X V_{2k+2}^{(h/2)}(z)P_{\frac{h}{2}}(\mathrm{d}z|y,a)P_{\frac{h}{2}}(\mathrm{d}y|x,a)-\int_X V_{2k+2}^{(h/2)}(y)P_h(\mathrm{d}y|x,a)\Big\} \tag{3.6}$$

for all $x\in X$. Moreover, direct calculations, together with (3.1), Lemma 3.3, Assumptions 2.1 and 3.1, yield

$$\begin{aligned}\Big|\int_X[c(y,a)-c(x,a)]P_{\frac{h}{2}}(\mathrm{d}y|x,a)\Big| &= [(h/2)m(x)\wedge 1]\Big|\frac{1}{m(x)}\int_X c(y,a)q(\mathrm{d}y|x,a)\Big|\\ &\le \frac{hM}{2}\Big[\int_X\omega(y)q(\mathrm{d}y|x,a)+2q^*(x)\omega(x)\Big]\\ &\le \frac{h}{2}M(b_1+2L)\omega^2(x),\end{aligned} \tag{3.7}$$

and

$$\begin{aligned}&\Big|\int_X\int_X V_{2k+2}^{(h/2)}(z)P_{\frac{h}{2}}(\mathrm{d}z|y,a)P_{\frac{h}{2}}(\mathrm{d}y|x,a)-\int_X V_{2k+2}^{(h/2)}(y)P_h(\mathrm{d}y|x,a)\Big|\\ =\ &\Big|\int_X\Big[\int_X\big(V_{2k+2}^{(h/2)}(z)-V_{2k+2}^{(h/2)}(y)\big)P_{\frac{h}{2}}(\mathrm{d}z|y,a)\Big]P_{\frac{h}{2}}(\mathrm{d}y|x,a)\\ &-\Big[\int_X V_{2k+2}^{(h/2)}(y)P_{\frac{h}{2}}(\mathrm{d}y|x,a)-V_{2k+2}^{(h/2)}(x)\Big]\\ &+2\Big[\int_X V_{2k+2}^{(h/2)}(y)P_{\frac{h}{2}}(\mathrm{d}y|x,a)-V_{2k+2}^{(h/2)}(x)\Big]\\ &-\Big[\int_X V_{2k+2}^{(h/2)}(y)P_h(\mathrm{d}y|x,a)-V_{2k+2}^{(h/2)}(x)\Big]\Big|\\ =\ &\Big|[(h/2)m(x)\wedge 1]\frac{1}{m(x)}\int_X\Big[\int_X\big(V_{2k+2}^{(h/2)}(z)-V_{2k+2}^{(h/2)}(y)\big)P_{\frac{h}{2}}(\mathrm{d}z|y,a)\Big]q(\mathrm{d}y|x,a)\\ &+\big(2[(h/2)m(x)\wedge 1]-[hm(x)\wedge 1]\big)\frac{1}{m(x)}\int_X V_{2k+2}^{(h/2)}(y)q(\mathrm{d}y|x,a)\Big|\\ =\ &\Big|[(h/2)m(x)\wedge 1]^2\frac{1}{m^2(x)}\int_X\Big[\int_X V_{2k+2}^{(h/2)}(z)q(\mathrm{d}z|y,a)\Big]q(\mathrm{d}y|x,a)\\ &+\big(2[(h/2)m(x)\wedge 1]-[hm(x)\wedge 1]\big)\frac{1}{m(x)}\int_X V_{2k+2}^{(h/2)}(y)q(\mathrm{d}y|x,a)\Big|\\ \le\ &\frac{h^2}{4}MTe^{b_1T}\int_X\Big[\int_X\omega(z)q(\mathrm{d}z|y,a)+2q^*(y)\omega(y)\Big]|q(\mathrm{d}y|x,a)|\\ &+h^2MTe^{b_1T}m(x)\Big[\int_X\omega(y)q(\mathrm{d}y|x,a)+2q^*(x)\omega(x)\Big]\\ \le\ &\frac{1}{4}MTe^{b_1T}\big(\rho_2+b_2+2L+4\|m\|_\omega\big)(b_1+2L)h^2\omega^3(x)\end{aligned} \tag{3.8}$$

for all $(x,a)\in K$, where the first inequality is due to the following fact that

$$2[(h/2)m(x)\wedge 1]-[hm(x)\wedge 1]=\begin{cases}0, & \text{if } hm(x)<1,\\ hm(x)-1, & \text{if } 1\le hm(x)<2,\\ 1, & \text{if } hm(x)\ge 2.\end{cases}$$

Hence, by (3.6) – (3.8) and the induction hypothesis, we have

$$\begin{aligned}
&V_{2k}^{(h/2)}(x)\\
\le\ &\inf_{a\in A(x)}\left\{hc(x,a)+\int_X V_{2k+2}^{(h/2)}(y)P_h(\mathrm{d}y|x,a)\right\}\\
&+\left[\frac{1}{4}M+\frac{1}{4}MTe^{b_1T}\big(\rho_2+b_2+2L+4\|m\|_\omega\big)\right](b_1+2L)h^2\omega^3(x)\\
\le\ &\inf_{a\in A(x)}\left\{hc(x,a)+\int_X V_{k+1}^h(y)P_h(\mathrm{d}y|x,a)+\left[\sum_{l=0}^{N-k-2}L^*h^2(1+\rho_3h+b_3h)^l\right.\right.\\
&\left.\left.+L^*h(1+\rho_3h+b_3h)^{N-k-1}\right]\int_X\omega^3(y)P_h(\mathrm{d}y|x,a)\right\}\\
&+\left[\frac{1}{4}M+\frac{1}{4}MTe^{b_1T}\big(\rho_2+b_2+2L+4\|m\|_\omega\big)\right](b_1+2L)h^2\omega^3(x)\\
\le\ &V_k^h(x)+\left[\sum_{l=0}^{N-k-2}L^*h^2(1+\rho_3h+b_3h)^{l+1}+L^*h(1+\rho_3h+b_3h)^{N-k}+L^*h^2\right]\omega^3(x)\\
=\ &V_k^h(x)+\left[\sum_{l=0}^{N-k-1}L^*h^2(1+\rho_3h+b_3h)^{l}+L^*h(1+\rho_3h+b_3h)^{N-k}\right]\omega^3(x)
\end{aligned}$$

for all $x\in X$, where the third inequality is due to the fact that

$$\begin{aligned}
\int_X\omega^3(y)P_h(\mathrm{d}y|x,a)&=[hm(x)\wedge 1]\left\{\frac{1}{m(x)}\int_X\omega^3(y)q(\mathrm{d}y|x,a)\right\}+\omega^3(x)\\
&\le(1+\rho_3h+b_3h)\omega^3(x)
\end{aligned}$$

for all $(x,a)\in K$. This completes the proof of the lemma. □

Lemma 3.5. Under Assumptions 2.1 and 3.1, there exists a measurable function V on $X\times[0,T]$ such that for all $x\in X$ and $t\in[0,T]$,

$$V_{\lfloor th^{-1}\rfloor}^h(x)\to V(x,t)\text{ as }h=2^{-k}\text{ and }k\to\infty.$$

Proof. Fix any $t\in[0,T]$. For each $n=0,1,\dots,\lfloor(T-t)h^{-1}\rfloor-1$, $x\in X$ and $\pi=\{f_k,k=0,1,\dots\}\in\widetilde{\Pi}_d$, we define

$$\overline{V}_n^h(x,\pi):=\widetilde{E}_{n,x}^\pi\left[\sum_{k=n}^{\lfloor(T-t)h^{-1}\rfloor-1}hc(x_k,f_k(x_k))\right]\quad\text{and}\quad\overline{V}_n^h(x):=\inf_{\pi\in\widetilde{\Pi}_d}\overline{V}_n^h(x,\pi).\tag{3.9}$$

Then under Assumptions 2.1 and 3.1, employing Theorem 2.3.8 in [1], we have that there exists a policy $\pi^* = \{f_k^*, k = 0, 1, \ldots\} \in \widetilde{\Pi}_d$ such that $\overline{V}_0^h(x) = \overline{V}_0^h(x, \pi^*)$ for all $x \in X$. Let $\widetilde{\pi} = \{\widetilde{f}_k, k = 0, 1, \ldots\} \in \widetilde{\Pi}_d$ be a policy satisfying $\widetilde{f}_k = f^*_{k-N+\lfloor (T-t)h^{-1} \rfloor}$ for all $k = N - \lfloor (T-t)h^{-1} \rfloor, \ldots, N-1$. Thus, we have

$$\overline{V}_0^h(x) = \overline{V}_0^h(x, \pi^*) = V^h_{N-\lfloor (T-t)h^{-1} \rfloor}(x, \widetilde{\pi}) \geq V^h_{N-\lfloor (T-t)h^{-1} \rfloor}(x) \tag{3.10}$$

for all $x \in X$, where the second equality follows from Tulcea theorem in [15, p.178]. On the other hand, by the similar arguments of (3.10), we get $\overline{V}_0^h(x) \leq V^h_{N-\lfloor (T-t)h^{-1} \rfloor}(x)$ for all $x \in X$. Hence, we have

$$\overline{V}_0^h(x) = V^h_{N-\lfloor (T-t)h^{-1} \rfloor}(x) \quad \text{for all } x \in X. \tag{3.11}$$

For simplicity, we write $\overline{V}_n^{2^{-k}}$ as $\overline{V}_n^k$. Then, by Lemma 3.4 and the inequality $1 + z \leq e^z$ for all $z > 0$, we have

$$\begin{aligned} \overline{V}_0^{(k+1)}(x) &\leq \overline{V}_0^k(x) + \Bigg[\sum_{l=0}^{N-1} 2^{-2k} L^* (1 + 2^{-k}\rho_3 + 2^{-k} b_3)^l \\ &\qquad + 2^{-k} L^* (1 + 2^{-k}\rho_3 + 2^{-k} b_3)^N \Bigg] \omega^3(x) \\ &\leq \overline{V}_0^k(x) + L^*(T+1) e^{(\rho_3 + b_3)T} 2^{-k} \omega^3(x) \end{aligned} \tag{3.12}$$

for all $x \in X$ and $k = 0, 1, \ldots$. Iterating (3.12), we obtain

$$\begin{aligned} \overline{V}_0^{(n+l)}(x) &\leq \overline{V}_0^n(x) + 2^{-n} L^*(T+1) e^{(\rho_3 + b_3)T} \sum_{j=0}^{l-1} 2^{-j} \omega^3(x) \\ &\leq \overline{V}_0^n(x) + 2^{-n+1} L^*(T+1) e^{(\rho_3 + b_3)T} \omega^3(x) \end{aligned}$$

for all $x \in X$, $n = 0, 1, \ldots$, and $l = 1, 2, \ldots$, which gives

$$\limsup_{l \to \infty} \overline{V}_0^l(x) \leq \overline{V}_0^n(x) + 2^{-n+1} L^*(T+1) e^{(\rho_3 + b_3)T} \omega^3(x) \tag{3.13}$$

for all $x \in X$, $n = 0, 1, \ldots$. Thus, by (3.13), we see that $\lim_{k \to \infty} \overline{V}_0^k(x)$ exists and denote

$$V(x, t) := \lim_{k \to \infty} \overline{V}_0^k(x) \quad \text{for all } x \in X. \tag{3.14}$$

Observe that

$$t - h < \lfloor t h^{-1} \rfloor h \leq t \quad \text{and} \quad t - h < (N - \lfloor (T-t)h^{-1} \rfloor)h < t + h,$$

together with (3.3) and (3.11), yield

$$V^h_{\lfloor th^{-1} \rfloor}(x) = \overline{V}_0^h(x), \quad \text{or} \quad V^h_{\lfloor th^{-1} \rfloor}(x) = \inf_{a \in A(x)} \left\{ hc(x, a) + \int_X \overline{V}_0^h(y) P_h(\mathrm{d}y | x, a) \right\} \tag{3.15}$$

for all $x \in X$. Hence, it follows from (3.15) and Lemma 3.3 that

$$\begin{aligned}\left|V^h_{\lfloor th^{-1}\rfloor}(x) - \overline{V}^h_0(x)\right| &\le \sup_{a\in A(x)}\left|hc(x,a) + \int_X \overline{V}^h_0(y)P_h(\mathrm{d}y|x,a) - \overline{V}^h_0(x)\right| \\ &= \sup_{a\in A(x)}\left|hc(x,a) + [hm(x)\wedge 1]\frac{1}{m(x)}\int_X \overline{V}^h_0(y)q(\mathrm{d}y|x,a)\right| \\ &\le [1 + Te^{b_1T}(b_1+2L)]Mh\omega^2(x) \qquad (3.16)\end{aligned}$$

for all $x \in X$. Therefore, the following inequality

$$\left|V^h_{\lfloor th^{-1}\rfloor}(x) - V(x,t)\right| \le \left|V^h_{\lfloor th^{-1}\rfloor}(x) - \overline{V}^h_0(x)\right| + \left|\overline{V}^h_0(x) - V(x,t)\right|$$

for all $x \in X$, together with (3.14) and (3.16), implies the desired result. □

Now we give the following lemma which is used to prove our main results.

Lemma 3.6. Under Assumptions 2.1 and 3.1(iii), the following assertions hold.

(a) For each $x \in X$ and $u \in B_\omega(X)$, $\int_X u(y)q(\mathrm{d}y|x,a)$ is continuous in $a \in A(x)$.

(b) Let $\{u_n : n \ge 1\}$ be a bounded sequence in $B_\omega(X)$ (i. e., there exists a constant $\overline{L} > 0$ such that $\|u_n\|_\omega \le \overline{L}$ for all $n \ge 1$), and $\lim_{n\to\infty} u_n = u$. Then, for any $x \in X$ and any sequence $\{a_n : n \ge 1\}$ in $A(x)$ such that $a_n \to a^*$ in $A(x)$, we have

$$\lim_{n\to\infty}\int_X u_n(y)q(\mathrm{d}y|x,a_n) = \int_X u(y)q(\mathrm{d}y|x,a^*).$$

P r o o f. (a) Let $\tilde{u} := u + \|u\|_\omega\omega$. Then, we have $\tilde{u} \ge 0$ and $\tilde{u} \in B_\omega(X)$. For each $k \ge 1$, define $u_k := \tilde{u} \wedge k$. Fix any $x \in X$. Let $\{a_n : n \ge 1\}$ be a sequence in $A(x)$ converging to $a \in A(x)$. By Assumption 3.1(iii), we have

$$q(x|x,a_n) = \int_X I_{\{x\}}(y)q(\mathrm{d}y|x,a_n) \to \int_X I_{\{x\}}(y)q(\mathrm{d}y|x,a) = q(x|x,a) \text{ as } n \to \infty.$$

For each $k \ge 1$, applying Assumption 3.1(iii) to u_k, we obtain

$$\begin{aligned}\liminf_{n\to\infty}\int_{X\setminus\{x\}}\tilde{u}(y)q(\mathrm{d}y|x,a_n) &\ge \liminf_{n\to\infty}\int_{X\setminus\{x\}}u_k(y)q(\mathrm{d}y|x,a_n) \\ &= \liminf_{n\to\infty}\left[\int_X u_k(y)q(\mathrm{d}y|x,a_n) - u_k(x)q(x|x,a_n)\right] \\ &= \int_{X\setminus\{x\}}u_k(y)q(\mathrm{d}y|x,a),\end{aligned}$$

which, together with the monotone convergence theorem in [15, p. 170], gives

$$\liminf_{n\to\infty}\int_{X\setminus\{x\}}\tilde{u}(y)q(\mathrm{d}y|x,a_n) \ge \int_{X\setminus\{x\}}\tilde{u}(y)q(\mathrm{d}y|x,a).$$

Note that $\lim_{n\to\infty} \tilde{u}(x)q(x|x,a_n) = \tilde{u}(x)q(x|x,a)$. Hence, we get

$$\liminf_{n\to\infty} \int_X \tilde{u}(y)q(\mathrm{d}y|x,a_n) \geq \int_X \tilde{u}(y)q(\mathrm{d}y|x,a).$$

Since $u = \tilde{u} - \|u\|_\omega \omega$ and $\int_X \omega(x)q(y|x,a)$ is continuous in $a \in A(x)$, we have

$$\liminf_{n\to\infty} \int_X u(y)q(\mathrm{d}y|x,a_n) \geq \int_X u(y)q(\mathrm{d}y|x,a).$$

Replacing u with $-u$, we obtain

$$\lim_{n\to\infty} \int_X u(y)q(\mathrm{d}y|x,a_n) = \int_X u(y)q(\mathrm{d}y|x,a),$$

and so part (a) holds.

(b) For each $k \geq 1$, applying part (a) to $U_k := \inf_{n\geq k} u_n \in B_\omega(X)$, we have

$$\begin{aligned}
\liminf_{n\to\infty} \int_{X\setminus\{x\}} u_n(y)q(\mathrm{d}y|x,a_n) &\geq \liminf_{n\to\infty} \int_{X\setminus\{x\}} U_k(y)q(\mathrm{d}y|x,a_n) \\
&= \liminf_{n\to\infty} \left[\int_X U_k(y)q(\mathrm{d}y|x,a_n) - U_k(x)q(x|x,a_n)\right] \\
&= \int_{X\setminus\{x\}} U_k(y)q(\mathrm{d}y|x,a^*).
\end{aligned}$$

Observe that $|U_k(x)| \leq \overline{L}\omega(x)$ for all $x \in X$ and $k \geq 1$. Thus, by Assumption 2.1, we obtain

$$\int_{X\setminus\{x\}} |U_k(y)|q(\mathrm{d}y|x,a) \leq \overline{L}\int_{X\setminus\{x\}} \omega(y)q(\mathrm{d}y|x,a) \leq \overline{L}(b_1+L)\omega^2(x)$$

for all $(x,a) \in K$, which, together with the dominated convergence theorem in [15, p. 171] and Assumption 3.1(iii), yields

$$\begin{aligned}
\liminf_{n\to\infty} \int_X u_n(y)q(\mathrm{d}y|x,a_n) &\geq \liminf_{n\to\infty} \int_{X\setminus\{x\}} u_n(y)q(\mathrm{d}y|x,a_n) + \liminf_{n\to\infty} u_n(x)q(x|x,a_n) \\
&\geq \lim_{k\to\infty} \int_{X\setminus\{x\}} U_k(y)q(\mathrm{d}y|x,a^*) + u(x)q(x|x,a^*) \\
&= \int_X u(y)q(\mathrm{d}y|x,a^*).
\end{aligned}$$

Using the similar arguments, we have

$$\limsup_{n\to\infty} \int_X u_n(y)q(\mathrm{d}y|x,a_n) \leq \int_X u(y)q(\mathrm{d}y|x,a^*).$$

Hence, we obtain the desired result. □

For any $s \in [0,T]$, a function u on $X \times [s,T]$ is said to be $[s,T]$-*uniformly* ω^2-*bounded* if it is $\mathcal{B}(X \times [s,T])$-measurable and there exists a constant $\widetilde{M} > 0$ such that $|u(x,t)| \le \widetilde{M}\omega^2(x)$ for all $(x,t) \in X \times [s,T]$.

Then we have the following lemma.

Lemma 3.7. For any $s \in [0,T]$, $z \in X$, and $\pi \in \Pi$, define

$$\begin{aligned}\mathcal{H}_{s,z} := \Bigg\{ u :\ & u \text{ is } [s,T]\text{-uniformly } \omega^2\text{-bounded and satisfies} \\ & \int_s^T \int_X \int_X \Big[-\int_t^T u(y,v)\,\mathrm{d}v \Big] q(\mathrm{d}y|x,\pi_t) p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t \\ & = \int_s^T u(z,t)\,\mathrm{d}t - \int_s^T \int_X u(y,t) p_\pi(s,z,t,\mathrm{d}y)\,\mathrm{d}t \Bigg\}.\end{aligned}$$

Then, under Assumptions 2.1 and 3.1, the set $\mathcal{H}_{s,z}$ contains all $[s,T]$-uniformly ω^2-bounded functions.

P r o o f . It follows from Assumptions 2.1, 3.1, and Theorem 3.1 in [12] that

$$\begin{aligned} & \int_s^T \int_X \int_X \omega^2(y) |q(\mathrm{d}y|x,\pi_t)| p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t \\ \le\ & \int_s^T \int_X [\rho_2 + b_2 + 2L]\omega^3(x) p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t \\ \le\ & \big[\rho_2 + b_2 + 2L\big] \int_s^T \Big[e^{\rho_3(t-s)}\omega^3(z) + \frac{b_3}{\rho_3}(e^{\rho_3(t-s)} - 1) \Big]\,\mathrm{d}t \\ \le\ & \big[\rho_2 + b_2 + 2L\big] \Big[e^{\rho_3(T-s)}\omega^3(z) + \frac{b_3}{\rho_3}(e^{\rho_3(T-s)} - 1) \Big](T-s). \end{aligned} \tag{3.17}$$

Hence, if u is $[s,T]$-uniformly ω^2-bounded, by (3.17), we have

$$\int_s^T \int_X \int_X \Big| -\int_t^T u(y,v)\,\mathrm{d}v \Big| |q(\mathrm{d}y|x,\pi_t)| p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t < \infty. \tag{3.18}$$

Therefore, $\int_s^T \int_X \int_X [-\int_t^T u(y,v)\,\mathrm{d}v] q(\mathrm{d}y|x,\pi_t) p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t$ is well defined. Similarly, we can show that $\int_s^T \int_X u(y,t) p_\pi(s,z,t,\mathrm{d}y)\,\mathrm{d}t$ is well defined. Define $\mathcal{C} := \{B \times [k,l] : B \in \mathcal{B}(X),\ s \le k \le l \le T\}$. Then, we see that $\mathcal{C}$ is a π-system and $X \times [s,T] \in \mathcal{C}$. Next, we will use the monotone class theorem to show that $\mathcal{H}_{s,z}$ contains all bounded $\mathcal{B}(X \times [s,T])$-measurable functions.

(i) For any $B \times [k,l] \in \mathcal{C}$, we will show that $I_B(y) I_{[k,l]}(t) \in \mathcal{H}_{s,z}$. By the Kolmogorov forward equation and direct calculations, we have

$$\int_s^T \int_X \int_X \Big[-\int_t^T I_B(y) I_{[k,l]}(v)\,\mathrm{d}v \Big] q(\mathrm{d}y|x,\pi_t) p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t$$

$$
\begin{aligned}
&= \int_s^T \int_X \int_X I_B(y)[(k \vee t) \wedge l - l] q(\mathrm{d}y|x, \pi_t) p_\pi(s, z, t, \mathrm{d}x)\, \mathrm{d}t \\
&= (k-l) \int_s^k \int_X q(B|x, \pi_t) p_\pi(s, z, t, \mathrm{d}x)\, \mathrm{d}t + \int_k^l \int_X (t-l) q(B|x, \pi_t) p_\pi(s, z, t, \mathrm{d}x)\, \mathrm{d}t \\
&= (k-l) \int_s^k \frac{\partial p_\pi(s, z, t, B)}{\partial t}\, \mathrm{d}t + \int_k^l (t-l) \frac{\partial p_\pi(s, z, t, B)}{\partial t}\, \mathrm{d}t \\
&= -(k-l) I_B(z) - \int_k^l p_\pi(s, z, t, B)\, \mathrm{d}t \\
&= \int_s^T I_B(z) I_{[k,l]}(t)\, \mathrm{d}t - \int_s^T \int_X I_B(y) I_{[k,l]}(t) p_\pi(s, z, t, \mathrm{d}y)\, \mathrm{d}t,
\end{aligned}
$$

where $y_1 \vee y_2 := \max\{y_1, y_2\}$. Hence, we have $I_B(y) I_{[k,l]}(t) \in \mathcal{H}_{s,z}$.

(ii) If $0 \le u_n \in \mathcal{H}_{s,z} (n = 1, 2, \ldots)$, $u_n \uparrow u_0$ and u_0 is bounded, we will show that $u_0 \in \mathcal{H}_{s,z}$. By the monotone convergence theorem in [15, p. 170], we see that

$$
\begin{aligned}
&\int_s^T u_n(z, t)\, \mathrm{d}t - \int_s^T \int_X u_n(y, t) p_\pi(s, z, t, \mathrm{d}y)\, \mathrm{d}t \\
\to &\int_s^T u_0(z, t)\, \mathrm{d}t - \int_s^T \int_X u_0(y, t) p_\pi(s, z, t, \mathrm{d}y)\, \mathrm{d}t
\end{aligned}
$$

as $n \to \infty$. Since every bounded $\mathcal{B}(X \times [s, T])$-measurable function is $[s, T]$-uniformly ω^2-bounded, u_0 satisfies (3.18). Thus, we have

$$
\int_s^T \int_X \int_{X \setminus \{x\}} \int_t^T u_0(y, v)\, \mathrm{d}v q(\mathrm{d}y|x, \pi_t) p_\pi(s, z, t, \mathrm{d}x)\, \mathrm{d}t < \infty, \quad \text{and}
$$

$$
\int_s^T \int_X \int_t^T u_0(x, v)\, \mathrm{d}v q(x|x, \pi_t) p_\pi(s, z, t, \mathrm{d}x)\, \mathrm{d}t < \infty.
$$

Hence, using the monotone convergence theorem, we obtain

$$
\begin{aligned}
&\int_s^T \int_X \int_X \Big[- \int_t^T u_n(y, v)\, \mathrm{d}v \Big] q(\mathrm{d}y|x, \pi_t) p_\pi(s, z, t, \mathrm{d}x)\, \mathrm{d}t \\
= &\int_s^T \int_X \int_{X \setminus \{x\}} \Big[- \int_t^T u_n(y, v)\, \mathrm{d}v \Big] q(\mathrm{d}y|x, \pi_t) p_\pi(s, z, t, \mathrm{d}x)\, \mathrm{d}t \\
&+ \int_s^T \int_X \Big[- \int_t^T u_n(x, v)\, \mathrm{d}v \Big] q(x|x, \pi_t) p_\pi(s, z, t, \mathrm{d}x)\, \mathrm{d}t \\
\to &\int_s^T \int_X \int_X \Big[- \int_t^T u_0(y, v)\, \mathrm{d}v \Big] q(\mathrm{d}y|x, \pi_t) p_\pi(s, z, t, \mathrm{d}x)\, \mathrm{d}t,
\end{aligned}
$$

as $n \to \infty$. From the above discussion, we get $u_0 \in \mathcal{H}_{s,z}$.

It is obvious that $\mathcal{H}_{s,z}$ is a linear space. Thus, it follows from (i), (ii), and the monotone class theorem that $\mathcal{H}_{s,z}$ contains all bounded $\mathcal{B}(X \times [s, T])$-measurable functions.

For any $[s,T]$-uniformly ω^2-bounded function u, we have $u = u^+ - u^-$, where $u^+ = u \vee 0$ and $u^- = (-u) \vee 0$. Since $\mathcal{H}_{s,z}$ is a linear space, without loss of generality, we may assume $u \geq 0$. For each $n \geq 1$, define $u_n := u \wedge n$. Then, u_n is a bounded $\mathcal{B}(X \times [s,T])$-measurable function, and so $u_n \in \mathcal{H}_{s,z}$ for each $n \geq 1$. Using the similar proof of (ii), we have $u \in \mathcal{H}_{s,z}$. Hence, $\mathcal{H}_{s,z}$ contains all $[s,T]$-uniformly ω^2-bounded functions. □

Finally, for ease of reference, we state the following lemma from [11], which is used to prove the existence of strong average optimal policies.

Lemma 3.8. Under Assumptions 2.1, 3.1, and 3.2, the following statements hold.

(a) There exist a constant g^*, two functions $u_1, u_2 \in B_\omega(X)$, and a stationary policy $f^* \in F$, satisfying the following two average optimality inequalities:

$$\begin{aligned} g^* &\leq \inf_{a \in A(x)} \left\{ c(x,a) + \int_X u_1(y) q(\mathrm{d}y|x,a) \right\}, \\ g^* &\geq \inf_{a \in A(x)} \left\{ c(x,a) + \int_X u_2(y) q(\mathrm{d}y|x,a) \right\} \qquad (3.19) \\ &= c(x, f^*(x)) + \int_X u_2(y) q(\mathrm{d}y|x, f^*(x)) \qquad (3.20) \end{aligned}$$

for all $x \in X$.

(b) $g^* = J^*(x) = J(x, f^*)$ for all $x \in X$.

Proof. See Theorem 4.2 in [11] for the proof. □

4. MAIN RESULTS

In this section, we state and prove our main results.

Now we present the result on the finite-horizon expected total cost criterion.

Theorem 4.1. Under Assumptions 2.1 and 3.1, the following statements hold.

(a) The finite-horizon optimal value function V_T^* on $X \times [0,T]$ is a solution to the following equation: for each $x \in X$ and $t \in [0,T]$,

$$u(x,t) = \int_t^T \inf_{a \in A(x)} \left\{ c(x,a) + \int_X u(y,s) q(\mathrm{d}y|x,a) \right\} \mathrm{d}s,$$

where the measurable function u on $X \times [0,T]$ satisfies $\sup_{x \in X} \sup_{t \in [0,T]} \frac{|u(x,t)|}{\omega(x)} < \infty$.

(b) There exists an optimal deterministic Markov policy π_T^* (depending on T).

Proof. (a) Fix any $t \in [0,T]$. For each $h > 0$, define the operators G and G^h on $B_\omega(X)$ as follows:

$$Gu(x) := \inf_{a \in A(x)} \left\{ c(x,a) + \int_X u(y) q(\mathrm{d}y|x,a) \right\}$$

$$G^h u(x) := h^{-1}\left[\inf_{a\in A(x)}\left\{hc(x,a)+\int_X u(y)P_h(\mathrm{d}y|x,a)\right\}-u(x)\right]$$

for all $u \in B_\omega(X)$ and $x \in X$. Then, by (3.3), we have

$$V_k^h(x) - V_{k+1}^h(x) = hG^h V_{k+1}^h(x)$$

for all $x \in X$, $h > 0$, and $k = 0, 1, \ldots, N-1$, which gives

$$\begin{aligned}
&V^h_{\lfloor th^{-1}\rfloor}(x)\\
&= \sum_{k=1}^{N-\lfloor th^{-1}\rfloor} hG^h V^h_{\lfloor th^{-1}\rfloor+k}(x)\\
&= \sum_{k=1}^{N-\lfloor th^{-1}\rfloor} hGV^h_{\lfloor th^{-1}\rfloor+k}(x) + \sum_{k=1}^{N-\lfloor th^{-1}\rfloor} h\big(G^h V^h_{\lfloor th^{-1}\rfloor+k}(x) - GV^h_{\lfloor th^{-1}\rfloor+k}(x)\big)\\
&= \int_{\lfloor th^{-1}\rfloor h}^{Nh} GV^h_{\lfloor sh^{-1}\rfloor+1}(x)\,\mathrm{d}s + \sum_{k=1}^{N-\lfloor th^{-1}\rfloor} h\big(G^h V^h_{\lfloor th^{-1}\rfloor+k}(x) - GV^h_{\lfloor th^{-1}\rfloor+k}(x)\big) \qquad (4.1)
\end{aligned}$$

for all $x \in X$ and $h > 0$. Moreover, we have

$$\begin{aligned}
&\big|G^h V^h_{\lfloor th^{-1}\rfloor+k}(x) - GV^h_{\lfloor th^{-1}\rfloor+k}(x)\big|\\
\le\ &\sup_{a\in A(x)}\Big|h^{-1}\int_X V^h_{\lfloor th^{-1}\rfloor+k}(y)P_h(\mathrm{d}y|x,a) - h^{-1}V^h_{\lfloor th^{-1}\rfloor+k}(x)\\
&-\int_X V^h_{\lfloor th^{-1}\rfloor+k}(y)q(\mathrm{d}y|x,a)\Big|\\
=\ &\sup_{a\in A(x)}\left|\left\{[m(x)\wedge h^{-1}]\frac{1}{m(x)}-1\right\}\int_X V^h_{\lfloor th^{-1}\rfloor+k}(y)q(\mathrm{d}y|x,a)\right|\\
\le\ &\left\{1-[m(x)\wedge h^{-1}]\frac{1}{m(x)}\right\}MTe^{b_1T}(b_1+2L)\omega^2(x)\\
\le\ &\left\{1-[m(x)\wedge h^{-1}]\frac{1}{m(x)}\right\}\frac{1}{m(x)}\|m\|_\omega MTe^{b_1T}(b_1+2L)\omega^3(x)\\
\le\ &\|m\|_\omega MTe^{b_1T}(b_1+2L)h\omega^3(x) \qquad (4.2)
\end{aligned}$$

for all $x \in X$, $h > 0$, and $k = 1, \ldots, N - \lfloor th^{-1}\rfloor$, where the first equality follows from (3.1), and the second and fourth inequalities are due to Lemma 3.3, Assumption 2.1, and the following fact that if $m(x) > h^{-1}$,

$$\left\{1-[m(x)\wedge h^{-1}]\frac{1}{m(x)}\right\}\frac{1}{m(x)} = \left(1-\frac{1}{hm(x)}\right)\frac{1}{m(x)} \le h;$$

if $m(x) \le h^{-1}$,

$$\left\{1-[m(x)\wedge h^{-1}]\frac{1}{m(x)}\right\}\frac{1}{m(x)} = 0 \le h.$$

Hence, by (4.2), we obtain

$$\begin{aligned}&\left|\sum_{k=1}^{N-\lfloor th^{-1}\rfloor} h\big(G^h V^h_{\lfloor th^{-1}\rfloor+k}(x) - GV^h_{\lfloor th^{-1}\rfloor+k}(x)\big)\right| \\ \le\ & (N-\lfloor th^{-1}\rfloor)h\|m\|_\omega MTe^{b_1T}(b_1+2L)h\omega^3(x) \\ \le\ & \|m\|_\omega MT^2e^{b_1T}(b_1+2L)h\omega^3(x)\end{aligned}$$

for all $x \in X$ and $h > 0$, which implies

$$\sum_{k=1}^{N-\lfloor th^{-1}\rfloor} h\big(G^h V^h_{\lfloor th^{-1}\rfloor+k}(x) - GV^h_{\lfloor th^{-1}\rfloor+k}(x)\big) \to 0 \ \text{ as } h = 2^{-l} \text{ and } l \to \infty \tag{4.3}$$

for all $x \in X$. Note that

$$V^h_{\lfloor sh^{-1}\rfloor+1}(x) = V^h_{\lfloor sh^{-1}\rfloor}(x) - hG^h V^h_{\lfloor sh^{-1}\rfloor+1}(x),$$

and

$$\begin{aligned}\big|hG^h V^h_{\lfloor sh^{-1}\rfloor+1}(x)\big| &\le \sup_{a\in A(x)}\left\{hM\omega(x) + \left|[hm(x)\wedge 1]\frac{1}{m(x)}\int_X V^h_{\lfloor sh^{-1}\rfloor+1}(y)q(\mathrm{d}y|x,a)\right|\right\} \\ &\le \big[Te^{b_1T}(b_1+2L)+1\big]Mh\omega^2(x)\end{aligned}$$

for all $x \in X$, $s \in [0,T]$, and $h > 0$, together with Lemma 3.5, yield

$$V^h_{\lfloor sh^{-1}\rfloor+1}(x) \to V(x,s) \ \text{ as } h = 2^{-l} \text{ and } l \to \infty. \tag{4.4}$$

Define

$$H_l(x) := \int_{-\infty}^{\infty} I_{[\lfloor 2^l t\rfloor 2^{-l}, N(l)2^{-l}]}(s) \inf_{a\in A(x)}\left\{c(x,a) + \int_X V^{2^{-l}}_{\lfloor 2^l s\rfloor+1}(y)q(\mathrm{d}y|x,a)\right\}\mathrm{d}s$$

for all $x \in X$ and $l = 0, 1, \ldots$, where $N(l) := \lfloor T2^l\rfloor$. Observe that

$$\begin{aligned}&\left|\inf_{a\in A(x)}\left\{c(x,a) + \int_X V^{2^{-l}}_{\lfloor 2^l s\rfloor+1}(y)q(\mathrm{d}y|x,a)\right\} - \inf_{a\in A(x)}\left\{c(x,a) + \int_X V(y,s)q(\mathrm{d}y|x,a)\right\}\right| \\ &\le \sup_{a\in A(x)}\left|\int_X V^{2^{-l}}_{\lfloor 2^l s\rfloor+1}(y)q(\mathrm{d}y|x,a) - \int_X V(y,s)q(\mathrm{d}y|x,a)\right|\end{aligned} \tag{4.5}$$

for all $x \in X$ and $s \in [0,T]$. Moreover, for each fixed $x \in X$, $s \in [0,T]$, and each $l \ge 1$, Assumption 3.1, Lemmas 3.3 and 3.6 give the existence of $a_l \in A(x)$ such that

$$\begin{aligned}&\sup_{a\in A(x)}\left|\int_X V^{2^{-l}}_{\lfloor 2^l s\rfloor+1}(y)q(\mathrm{d}y|x,a) - \int_X V(y,s)q(\mathrm{d}y|x,a)\right| \\ =\ & \left|\int_X V^{2^{-l}}_{\lfloor 2^l s\rfloor+1}(y)q(\mathrm{d}y|x,a_l) - \int_X V(y,s)q(\mathrm{d}y|x,a_l)\right|.\end{aligned}$$

Next, we will show that for each $x \in X$ and $s \in [0,T]$,

$$\left| \int_X V^{2^{-l}}_{\lfloor 2^l s \rfloor + 1}(y) q(\mathrm{d}y|x,a_l) - \int_X V(y,s) q(\mathrm{d}y|x,a_l) \right| \to 0 \tag{4.6}$$

as $l \to \infty$. Suppose that (4.6) is not true. Then, there exist a constant $\varepsilon > 0$ and a subsequence $\{l_i\}$ of $\{l\}$ such that

$$\left| \int_X V^{2^{-l_i}}_{\lfloor 2^{l_i} s \rfloor + 1}(y) q(\mathrm{d}y|x,a_{l_i}) - \int_X V(y,s) q(\mathrm{d}y|x,a_{l_i}) \right| \geq \varepsilon \tag{4.7}$$

for all $i \geq 1$. Since $A(x)$ is compact, there exists a subsequence of $\{l_i\}$ (still denoted by $\{l_i\}$) such that $\lim\limits_{i\to\infty} a_{l_i} =: \overline{a}$ for some $\overline{a} \in A(x)$. Thus, by Lemmas 3.3, 3.6, and (4.4), we get

$$\left| \int_X V^{2^{-l_i}}_{\lfloor 2^{l_i} s \rfloor + 1}(y) q(\mathrm{d}y|x,a_{l_i}) - \int_X V(y,s) q(\mathrm{d}y|x,a_{l_i}) \right| \to 0$$

as $i \to \infty$, which contradicts (4.7), and so (4.6) holds. Hence, for each $x \in X$, the dominated convergence theorem in [15, p. 171], Lemma 3.5, (4.1), (4.3), (4.5) and (4.6) give

$$V(x,t) = \lim_{l\to\infty} H_l(x) = \int_t^T \inf_{a\in A(x)} \left\{ c(x,a) + \int_X V(y,s) q(\mathrm{d}y|x,a) \right\} \mathrm{d}s. \tag{4.8}$$

Therefore, for each $x \in X$, the function $V(x,\cdot)$ is absolutely continuous on $[0,T]$. Using the similar proof of (4.6), we can show that for each $x \in X$, the function

$$\inf_{a\in A(x)} \left\{ c(x,a) + \int_X V(y,s) q(\mathrm{d}y|x,a) \right\}$$

is continuous in $s \in [0,T]$, which, together with (4.8), yields that the partial derivative of V with respect to the second variable t exists, denoted by $\frac{\partial V}{\partial t}$. By Assumptions 2.1, 3.1, and Theorem 3.1 in [12], we obtain

$$\begin{aligned} \left| E^\pi_{s,x} \left[\int_s^T \int_A c(x(t),a) \pi_t(\mathrm{d}a|x(t)) \, \mathrm{d}t \right] \right| &\leq M \int_s^T \left[e^{\rho_1(t-s)} \omega(x) + \frac{b_1}{\rho_1} \big(e^{\rho_1(t-s)} - 1 \big) \right] \mathrm{d}t \\ &\leq MT \left[e^{\rho_1 T} + \frac{b_1}{\rho_1} (e^{\rho_1 T} - 1) \right] \omega(x) \end{aligned}$$

for all $\pi \in \Pi$, $x \in X$, and $s \in [0,T]$, which gives

$$\sup_{x\in X} \sup_{t\in[0,T]} \frac{|V^*_T(x,t)|}{\omega(x)} \leq MT \left[e^{\rho_1 T} + \frac{b_1}{\rho_1} (e^{\rho_1 T} - 1) \right] < \infty.$$

By (4.8), we have

$$-\frac{\partial V}{\partial t}(x,t) = \inf_{a\in A(x)} \left\{ c(x,a) + \int_X V(y,t) q(\mathrm{d}y|x,a) \right\} \tag{4.9}$$

for all $x \in X$ and $t \in [0,T]$. Then, by (4.9), we obtain

$$-\frac{\partial V}{\partial t}(x,t) \le c(x,\pi_t) + \int_X V(y,t)q(\mathrm{d}y|x,\pi_t) \tag{4.10}$$

for all $\pi \in \Pi$, $t \in [0,T]$, and $x \in X$, where $c(x,\pi_t) := \int_{A(x)} c(x,a)\pi_t(\mathrm{d}a|x)$. Since $V(x,T) = 0$ for all $x \in X$, we have $V(x,t) = -\int_t^T \frac{\partial V}{\partial v}(x,v)\,\mathrm{d}v$ for each $t \in [0,T]$. Hence, it follows from (4.10), Fubini theorem, and Theorem 2.5 in [13, p. 15] that for each $\pi \in \Pi$, $z \in X$, and $s \in [0,T]$,

$$\begin{aligned}
&-\int_s^T \int_X \frac{\partial V}{\partial t}(x,t)p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t \\
\le\ & \int_s^T \int_X c(x,\pi_t)p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t + \int_s^T \int_X \int_X V(y,t)q(\mathrm{d}y|x,\pi_t)p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t \\
=\ & V_{T-s}(z,\pi) + \int_s^T \int_X \int_X \left[-\int_t^T \frac{\partial V}{\partial v}(y,v)\,\mathrm{d}v\right] q(\mathrm{d}y|x,\pi_t)p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t. \quad (4.11)
\end{aligned}$$

Moreover, it follows from Assumptions 2.1, 3.1, Lemmas 3.3 and 3.5 that

$$\begin{aligned}
\left|c(x,a) + \int_X V(y,t)q(\mathrm{d}y|x,a)\right| &\le M\omega(x) + MTe^{b_1T}\int_X \omega(y)|q(\mathrm{d}y|x,a)| \\
&\le [M + MTe^{b_1T}(b_1+2L)]\omega^2(x) \qquad (4.12)
\end{aligned}$$

for all $(x,a) \in K$ and $t \in [0,T]$. Using (4.9), (4.12), Assumption 3.1, and Theorem 3.1 in [12], we see that for each $\pi \in \Pi$, $z \in X$, and $s \in [0,T]$,

$$\begin{aligned}
&\left|\int_s^T \int_X \frac{\partial V}{\partial t}(x,t)p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t\right| \\
\le\ & [M + MTe^{b_1T}(b_1+2L)]\int_s^T \int_X \omega^2(x)p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t \\
\le\ & [M + MTe^{b_1T}(b_1+2L)]\int_s^T \left[e^{\rho_2(t-s)}\omega^2(z) + \frac{b_2}{\rho_2}(e^{\rho_2(t-s)}-1)\right]\mathrm{d}t \\
\le\ & [MT + MT^2e^{b_1T}(b_1+2L)]\left[e^{\rho_2T} + \frac{b_2}{\rho_2}\left(e^{\rho_2T}-1\right)\right]\omega^2(z).
\end{aligned}$$

From (4.12), we have that $\frac{\partial V}{\partial t}$ is a $[s,T]$-uniformly ω^2-bounded function. Thus, by Lemma 3.7, for each $\pi \in \Pi$, $z \in X$, and $s \in [0,T]$, we obtain

$$\begin{aligned}
&\int_s^T \int_X \int_X \left[-\int_t^T \frac{\partial V}{\partial v}(y,v)\,\mathrm{d}v\right] q(\mathrm{d}y|x,\pi_t)p_\pi(s,z,t,\mathrm{d}x)\,\mathrm{d}t \\
=\ & -V(z,s) - \int_s^T \int_X \frac{\partial V}{\partial t}(y,t)p_\pi(s,z,t,\mathrm{d}y)\,\mathrm{d}t,
\end{aligned}$$

which, together with (4.11), yields $V(z,s) \le V_{T-s}(z,\pi)$. Since π is arbitrary, we have

$$V(z,s) \le V_T^*(z,s) \ \text{ for all } z \in X \text{ and } s \in [0,T]. \tag{4.13}$$

On the other hand, Assumption 3.1 and the measurable selection theorem in [16, p. 50] give the existence of a Borel-measurable function f^* on $[0,T] \times X$ satisfying $f^*(t,x) \in A(x)$ and

$$-\frac{\partial V}{\partial t}(x,t) = c(x, f^*(t,x)) + \int_X V(y,t) q(\mathrm{d}y|x, f^*(t,x))$$

for all $x \in X$ and $t \in [0,T]$. For a policy $\pi^* = \{\pi_t^*, t \geq 0\} \in \Pi_d$ with $\pi_t^*(\cdot|x) = \delta_{f^*(t,x)}(\cdot)$ for all $x \in X$ and $t \in [0,T]$, where $\delta_a(\cdot)$ is the Dirac measure at $a \in A$, following the arguments of (4.13), we obtain

$$V(z,s) = V_{T-s}(z,\pi^*) \geq V_T^*(z,s) \quad \text{for all } z \in X \text{ and } s \in [0,T]. \tag{4.14}$$

Therefore, by (4.13) and (4.14), we have $V(z,s) = V_T^*(z,s)$ for all $z \in X$ and $s \in [0,T]$.

(b) The existence of an optimal deterministic Markov policy π_T^* follows obviously from (4.13) and (4.14). □

Remark 4.2. The optimality equation for finite-horizon expected total cost criterion has been established in [18] for finite states and finite actions, in [1] for bounded transition rates and denumerable state spaces, in [9, 10, 19, 22] for bounded transition rates and Borel state spaces, and in [4] for unbounded transition rates and denumerable state spaces. Theorem 4.1 extends the optimality equation in the aforementioned works to the case of *unbounded* transition rates and Borel spaces. It should be mentioned that the existence of optimal policies and the result that the finite-horizon optimal value function is a solution to the optimality equation have not been discussed in [4]. Moreover, the uniformization technique is inapplicable to the case of *unbounded* transition rates.

Next, we give the result on the existence of a strong average optimal policy.

Theorem 4.3. Under Assumptions 2.1, 3.1, and 3.2, the following statements hold.

(a) Every average optimal policy is strong average optimal.

(b) Any stationary policy $f \in F$ that attains the minimum of (3.19) is strong average optimal, and so f^* in (3.20) is a strong average optimal policy.

P r o o f. (a) We will first show the following equality

$$\lim_{T\to\infty} \frac{1}{T} V_T^*(x,0) = g^* \quad \text{for all } x \in X, \tag{4.15}$$

with g^* as in Lemma 3.8. On one hand, it follows from Lemma 3.8 that for each $\pi \in \Pi$, $z \in X$, and $t \in [0,T]$,

$$g^* \leq c(x,\pi_t) + \int_X u_1(y) q(\mathrm{d}y|x,\pi_t),$$

which, together with Fubini theorem and Theorem 2.5 in [13, p. 15], yields that

$$g^* \leq \frac{1}{T}\left[\int_0^T \int_X c(x,\pi_t) p_\pi(0,z,t,\mathrm{d}x)\,\mathrm{d}t + \int_0^T \int_X \int_X u_1(y) q(\mathrm{d}y|x,\pi_t) p_\pi(0,z,t,\mathrm{d}x)\,\mathrm{d}t\right]$$

$$= \frac{1}{T}V_T(z,\pi) + \frac{1}{T}\int_0^T \int_X \int_X u_1(y)q(\mathrm{d}y|x,\pi_t)p_\pi(0,z,t,\mathrm{d}x)\,\mathrm{d}t. \tag{4.16}$$

For each $z \in X$ and $\pi \in \Pi$, define

$$\mathcal{L}_z := \Big\{ u \in B_{\omega^2}(X) : \ \frac{1}{T}\int_0^T \int_X \int_X u(y)q(\mathrm{d}y|x,\pi_t)p_\pi(0,z,t,\mathrm{d}x)\,\mathrm{d}t = -\frac{1}{T}u(z) + \frac{1}{T}E_z^\pi[u(x(T))] \Big\}.$$

As in the proof of Lemma 3.7, we can show that $\mathcal{L}_z = B_{\omega^2}(X)$. Since $u_1 \in B_\omega(X) \subset B_{\omega^2}(X)$, by (4.16), we have

$$g^* \le \frac{1}{T}V_T(z,\pi) - \frac{1}{T}u_1(z) + \frac{1}{T}E_z^\pi[u_1(x(T))] \tag{4.17}$$

for all $\pi \in \Pi$ and $z \in X$. Note that

$$0 \le \lim_{T\to\infty} \frac{1}{T}\big|E_z^\pi[u(x(T))]\big| \le \lim_{T\to\infty} \frac{1}{T}\|u\|_\omega \left[e^{-\rho_1 T}\omega(x) + \frac{b_1}{\rho_1}(1 - e^{-\rho_1 T})\right] = 0 \tag{4.18}$$

for all $u \in B_\omega(X)$, $\pi \in \Pi$, and $z \in X$. Thus, letting $T \to \infty$ in (4.17), by (4.18), we get

$$\liminf_{T\to\infty} \frac{1}{T}V_T^*(z,0) = \liminf_{T\to\infty} \frac{1}{T}V_T(z,\pi_T^*) \ge g^* \tag{4.19}$$

for all $z \in X$, where π_T^* is as in Theorem 4.1. On the other hand, following the arguments of (4.17), by Lemma 3.8, we obtain

$$g^* \ge \frac{1}{T}V_T(z,f^*) + \frac{1}{T}E_z^{f^*}[u_2(x(T))] - \frac{1}{T}u_2(z)$$

for all $z \in X$, with $f^* \in F$ as in Lemma 3.8, which, together with (4.18), gives

$$g^* \ge \limsup_{T\to\infty} \frac{1}{T}V_T(z,f^*) \ge \limsup_{T\to\infty} \frac{1}{T}V_T^*(z,0) \tag{4.20}$$

for all $z \in X$. Hence, (4.15) follows from (4.19) and (4.20). By Lemma 3.8, we see that the set of average optimal policies is nonempty. Suppose that $\overline{\pi}$ is an arbitrary average optimal policy. Then, using Lemma 3.8, we have

$$\limsup_{T\to\infty} \frac{1}{T}V_T(x,\overline{\pi}) = J(x,\overline{\pi}) = J^*(x) = g^* \tag{4.21}$$

for all $x \in X$. Moreover, it follows from (4.15) that

$$g^* = \liminf_{T\to\infty} \frac{1}{T}V_T^*(x,0) \le \liminf_{T\to\infty} \frac{1}{T}V_T(x,\overline{\pi}) \tag{4.22}$$

for all $x \in X$. Thus, by (4.21) and (4.22), we have

$$\lim_{T\to\infty} \frac{1}{T}V_T(x,\overline{\pi}) = g^* \ \text{ for all } x \in X,$$

which, together with (4.15), implies that $\overline{\pi}$ is a strong average optimal policy.

(b) The proof of part (b) follows from part (a) and Lemma 3.8. □

Remark 4.4. Theorem 4.3 and Remark 2.3 indicate that the set of all average optimal policies coincides with the set of all strong average optimal policies, which is *new* for CTMDPs.

Finally, to verify Assumption 3.2, we provide a *new* set of sufficient conditions below.

Theorem 4.5. Under Assumptions 2.1 and 3.1, each of the following two sets of conditions implies Assumption 3.2.

(a) (The uniform ω-exponential ergodicity condition.) For each $f \in F$, there exists a probability measure μ_f on $\mathcal{B}(X)$ such that, for all $u \in B_\omega(X)$, $x \in X$, and $t \geq 0$,

$$\left| E_x^f[u(x(t))] - \int_X u(y)\mu_f(\mathrm{d}y) \right| \leq \|u\|_\omega R e^{-\eta t}\omega(x),$$

where the positive constants R, η are independent of f.

(b) (b1) The set $C := \{x \in X : \omega(x) \leq \frac{2b_1}{\rho_1}\}$ is nonempty.

(b2) There exist a constant $\xi > 0$ and a probability measure ν concentrated on the Borel set C such that $q(D \setminus \{x\}|x,a) + I_D(x) \geq \xi\nu(D)$ for each $D \in \mathcal{B}(C)$, $x \in C$, and $a \in A(x)$.

P r o o f. (a) Part (a) follows from Lemma 3.3 in [11].

(b) For each $x \in X$, $t \geq 0$, Borel set $D \subset C$, and $f \in F$, by Theorem 2 in [6], we have

$$\begin{aligned}
& p_f(0,x,t,D) \\
&= \int_0^t e^{q(x|x,f(x))z} \int_{X-\{x\}} p_f(z,y,t,D) q(\mathrm{d}y|x,f(x))\,\mathrm{d}z + I_D(x) e^{q(x|x,f(x))t} \\
&= \int_{X\setminus\{x\}} \left[\int_0^t e^{q(x|x,f(x))z} p_f(0,y,t-z,D)\,\mathrm{d}z \right] q(\mathrm{d}y|x,f(x)) + I_D(x) e^{q(x|x,f(x))t},
\end{aligned}$$

which, together with Assumption 2.1(ii) and condition (b1), gives

$$\begin{aligned}
& p_f(0,x,t,D) \\
&\geq \int_{X\setminus\{x\}} \left[\int_0^t e^{q(x|x,f(x))z} I_D(y) e^{q(y|y,f(y))(t-z)}\,\mathrm{d}z \right] q(\mathrm{d}y|x,f(x)) + e^{-L\omega(x)t} I_D(x) \\
&\geq \int_{D\setminus\{x\}} \left[\int_0^t e^{-L\omega(x)z} e^{-L\omega(y)t}\,\mathrm{d}z \right] q(\mathrm{d}y|x,f(x)) + e^{-L\omega(x)t} I_D(x) \\
&\geq t e^{-L\omega(x)t} \int_{D\setminus\{x\}} e^{-L\omega(y)t} q(\mathrm{d}y|x,f(x)) + e^{-L\omega(x)t} I_D(x) \\
&\geq t e^{-(L\omega(x)+\frac{2Lb_1}{\rho_1})t} q(D\setminus\{x\}|x,f(x)) + e^{-L\omega(x)t} I_D(x). \qquad (4.23)
\end{aligned}$$

Thus, for each $x \in C$, $t \geq 1$, $D \in \mathcal{B}(C)$ and $f \in F$, by (4.23), conditions (b1) and (b2), we obtain

$$p_f(0,x,t,D) \geq t e^{-\frac{4Lb_1 t}{\rho_1}} q(D\setminus\{x\}|x,f(x)) + e^{-\frac{2Lb_1 t}{\rho_1}} I_D(x)$$

$$\begin{aligned} &\geq e^{-\frac{4Lb_1t}{\rho_1}}\left(q(D\setminus\{x\}|x,f(x))+I_D(x)\right)\\ &\geq \xi e^{-\frac{4Lb_1t}{\rho_1}}\nu(D). \end{aligned} \tag{4.24}$$

Note that there exists $t_0 \geq 1$ such that $\kappa := \xi e^{-\frac{4Lb_1t_0}{\rho_1}} \in (0,1)$. Hence, for each $x \in C$, $D \in \mathcal{B}(C)$ and $f \in F$, it follows from (4.24) that

$$p_f(0,x,t_0,D) \geq \kappa\nu(D). \tag{4.25}$$

On the other hand, by Theorem 3.1 in [12] and Assumption 2.1(i), we have

$$\int_X \omega(y)p_f(0,x,t,\mathrm{d}y) \leq e^{-\rho_1 t}\omega(x) + \frac{b_1}{\rho_1}(1-e^{-\rho_1 t}), \tag{4.26}$$

which together with condition (b1) yields

$$\int_X \omega(y)p_f(0,x,t,\mathrm{d}y) \leq \frac{1}{2}(1+e^{-\rho_1 t})\omega(x) + \frac{b_1}{\rho_1}(1-e^{-\rho_1 t})I_C(x) \tag{4.27}$$

for all $f \in F$, $t \geq 0$, and $x \in X$. Therefore, by condition (b1), (4.25), (4.27), and Theorem 2.3 in [17], we see that for each $f \in F$, the t_0-skeleton chain $x^f_{t_0} := \{x(kt_0)|k = 0,1,2,\ldots\}$ with the one-step transition probability $Q(D|x,f(x)) := p_f(0,x,t_0,D)$ for all $x \in X$ and $D \in \mathcal{B}(X)$, is uniformly ω-geometrically ergodic. Thus, for each $f \in F$, there exist a probability measure μ_f on $\mathcal{B}(X)$, positive constants R_1 and $\eta_1 < 1$ (independent of f), such that

$$\left|\int_X u(y)p_f(0,x,nt_0,\mathrm{d}y) - \int_X u(y)\mu_f(\mathrm{d}y)\right| \leq \|u\|_\omega R_1\eta_1^n\omega(x) \tag{4.28}$$

for all $u \in B_\omega(X)$, $x \in X$, and $n = 0,1,2,\ldots$. Notice that for each $t > 0$, we have $t = lt_0 + s$ for some nonnegative integer l and $s \in [0,t_0)$. Hence, direct calculations, together with Chapman–Kolmogorov equation, Fubini theorem, (4.26), and (4.28), yield

$$\begin{aligned} &\left|E^f_x[u(x(t))] - \int_X u(y)\mu_f(\mathrm{d}y)\right|\\ =\ &\left|\int_X u(y)p_f(0,x,t,\mathrm{d}y) - \int_X u(y)\mu_f(\mathrm{d}y)\right|\\ =\ &\left|\int_X\int_X u(y)p_f(0,z,lt_0,\mathrm{d}y)p_f(0,x,s,\mathrm{d}z) - \int_X u(y)\mu_f(\mathrm{d}y)\right|\\ \leq\ &\int_X\left|\int_X u(y)p_f(0,z,lt_0,\mathrm{d}y) - \int_X u(y)\mu_f(\mathrm{d}y)\right|p_f(0,x,s,\mathrm{d}z)\\ \leq\ &\|u\|_\omega R_1\eta_1^l\int_X \omega(z)p_f(0,x,s,\mathrm{d}z)\\ \leq\ &\|u\|_\omega R_1\eta_1^l\left[e^{-\rho_1 s}\omega(x) + \frac{b_1}{\rho_1}(1-e^{-\rho_1 s})\right]\\ \leq\ &\|u\|_\omega R_1\eta_1^{-1}(\eta_1^{1/t_0})^t(1+b_1/\rho_1)\omega(x) \end{aligned}$$

for all $u \in B_\omega(X)$, and so condition (a) holds with $R := R_1\eta_1^{-1}(1 + b_1/\rho_1)$ and $\eta := -(\ln \eta_1)/t_0$. Hence, it follows from part (a) that condition (b) implies Assumption 3.2. □

Remark 4.6. (a) Theorem 4.5(a) has been established in [11] and indicates that Assumption 3.2 is weaker than the uniform ω-exponential ergodicity condition.

(b) The set of *verifiable* sufficient conditions imposed on the primitive data of the model for the verification of Assumption 3.2 in Theorem 4.5(b) is *new* and applicable to the case of denumerable state spaces. In particular, when X is a finite set, we usually choose $\omega = 1$, $b_1 \geq \rho_1$, and the set C in condition (b1) is equal to X.

5. AN EXAMPLE

In this section, a control problem in [14] is used to illustrate our results.

Example 5.1. The control model is given as follows: $X := (-\infty, \infty)$, $A = A(x) := [\theta_1, \theta_2]$ for all $x \in X$, with given positive constants $\theta_2 > \theta_1$, $q(D|x, a) := \beta\delta_0(D) + (\gamma|x| + a)\int_D \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy - (\gamma|x| + \beta + a)\delta_x(D)$ for all $D \in \mathcal{B}(X)$ and $(x, a) \in K$, where β and γ are given positive constants, and $\delta_x(\cdot)$ denotes the Dirac measure at $x \in X$.

To ensure the existence of a strong average optimal policy, we need the following hypotheses.

(C1) For each $x \in X$, $c(x, \cdot)$ is continuous on $A(x)$.

(C2) There exists a constant $\overline{M} > 0$ such that $|c(x, a)| \leq \overline{M}(x^2 + 1)$ for all $(x, a) \in K$.

Then we have the following result.

Proposition 5.2. Under conditions (C1) and (C2), Example 5.1 satisfies Assumptions 2.1, 3.1, and 3.2. Hence, (by Theorem 4.3), there exists a strong average optimal stationary policy for Example 5.1.

P r o o f . We first verify Assumption 2.1. Let $\rho_1 := \beta$, $b_1 := 2(\theta_2+\gamma)+\beta$, $L := 2\beta+\gamma+\theta_2$, $\omega(x) := x^2 + 1$ for all $x \in X$. Then, by the description of the model, a direct calculation yields

$$\int_X \omega(y)q(\mathrm{d}y|x, a) = \gamma|x| - (\gamma|x| + \beta)x^2 + a(1 - x^2) \tag{5.1}$$

for all $(x, a) \in K$. If $|x| \leq 1$, we have

$$\begin{aligned} \int_X \omega(y)q(\mathrm{d}y|x, a) &\leq \gamma|x| - (\gamma|x| + \beta)x^2 + \theta_2(1 - x^2) \\ &= \gamma(1 - x^2)|x| - (\beta + \theta_2)x^2 + \theta_2 \\ &\leq -(\beta + \gamma + \theta_2)\omega(x) + 2(\theta_2 + \gamma) + \beta \end{aligned} \tag{5.2}$$

for all $(x,a) \in K$. If $|x| > 1$, we have

$$\int_X \omega(y) q(\mathrm{d}y|x,a) \le \gamma(1-x^2)|x| - \beta x^2 \le -\beta\omega(x) + \beta \quad \text{for all } (x,a) \in K. \tag{5.3}$$

Thus, it follows from (5.2) and (5.3) that

$$\int_X \omega(y) q(\mathrm{d}y|x,a) \le -\beta\omega(x) + 2(\theta_2 + \gamma) + \beta \quad \text{for all } (x,a) \in K.$$

Moreover, by the definition of the transition rates, we obtain

$$q(\{x\}|x,a) = \beta\delta_0(\{x\}) - (\gamma|x| + \beta + a) \quad \text{for all } (x,a) \in K,$$

which gives $q^*(x) \le L\omega(x)$ for all $x \in X$. Hence, Assumption 2.1 is satisfied.

Now we verify Assumption 3.1. Let $\rho_2 := \beta + 6(\theta_2 + \gamma)$, $\rho_3 := \beta + 28(\theta_2 + \gamma)$, $b_2 = b_3 := 0$. Then we have

$$\begin{aligned} \int_X \omega^2(y) q(\mathrm{d}y|x,a) &= \beta + 6(\gamma|x| + a) - (\gamma|x| + \beta + a)(x^2+1)^2 \\ &\le [\beta + 6(\theta_2 + \gamma)]\omega^2(x), \end{aligned} \tag{5.4}$$

and

$$\begin{aligned} \int_X \omega^3(y) q(\mathrm{d}y|x,a) &= \beta + 28(\gamma|x| + a) - (\gamma|x| + \beta + a)(x^2+1)^3 \\ &\le [\beta + 28(\theta_2 + \gamma)]\omega^3(x), \end{aligned} \tag{5.5}$$

for all $(x,a) \in K$. Moreover, for each bounded measurable function u on X, a direct calculation gives

$$\int_X u(y) q(\mathrm{d}y|x,a) = \beta u(0) + (\gamma|x| + a) \int_X u(y) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \,\mathrm{d}y - (\gamma|x| + \beta + a)u(x)$$

for all $(x,a) \in K$, which implies that for each fixed $x \in X$, the function $\int_X u(y) q(\mathrm{d}y|x,a)$ is continuous in $a \in A(x)$. Hence, Assumption 3.1 follows from (5.1), (5.4), (5.5), and conditions (C1) and (C2).

Finally, we verify Assumption 3.2. Direct calculations yield

$$C := \left\{ x \in X : \omega(x) \le \frac{2b_1}{\rho_1} \right\} = \left[-\sqrt{\frac{4\gamma + 4\theta_2 + \beta}{\beta}}, \sqrt{\frac{4\gamma + 4\theta_2 + \beta}{\beta}} \right]. \tag{5.6}$$

Thus, the set C is nonempty. For each $D \in \mathcal{B}(C)$, $x \in C$ and $a \in A(x)$, by the description of the model, we have

$$\begin{aligned} q(D \setminus \{x\}|x,a) + I_D(x) &= \beta\delta_0(D \setminus \{x\}) + (\gamma|x| + a) \int_{D\setminus\{x\}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \,\mathrm{d}y + I_D(x) \\ &\ge \theta_1 \int_D \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \,\mathrm{d}y. \end{aligned}$$

Let $\xi := \theta_1 \int_C \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\,\mathrm{d}y$ and $\nu(D) := (\int_D \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\,\mathrm{d}y)(\int_C \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\mathrm{d}y)^{-1}$ for all $D \in \mathcal{B}(C)$. Hence, we obtain $q(D \setminus \{x\}|x,a) + I_D(x) \geq \xi\nu(D)$ for each $D \in \mathcal{B}(C)$, $x \in C$ and $a \in A(x)$. Therefore, by Theorem 4.5, we see that Assumption 3.2 holds. This completes the proof of the proposition. □

Remark 5.3. (a) The conditions in Example 5.1 above are *weaker* than those used in [14] since we have removed the following two hypotheses required in [14]: (i) the nonnegativity condition on the cost function, (ii) the condition "$\beta > \theta_2 + \frac{1}{2}\gamma$".

(b) The technique of the verification of the assumption that the relative difference of the discount optimal value function is bounded by an integrable function is *different* from that used in [14].

ACKNOWLEDGEMENT

The research was supported by the Fundamental Research Funds for the Central Universities of Huaqiao University (No.13SKGC-QT04). We are greatly indebted to the anonymous referee and the associate editor for many valuable comments and suggestions that have greatly improved the presentation.

(Received January 2, 2014)

REFERENCES

[1] N. Bäuerle and U. Rieder: Markov Decision Processes with Applications to Finance. Springer, Berlin 2011.

[2] D. P. Bertsekas and S. E. Shreve: Stochastic Optimal Control: The Discrete-time Case. Academic Press, New York 1978.

[3] R. Cavazos-Cadena and E. Fernández-Gaucherand: Denumerable controlled Markov chains with strong average optimality criterion: bounded and unbounded costs. Math. Methods Oper. Res. *43* (1996), 281–300.

[4] N. M. van Dijk: On the finite horizon Bellman equation for controlled Markov jump models with unbounded characteristics: existence and approximation. Stochastic Process. Appl. *28* (1988), 141–157.

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

[6] W. Feller: On the integro-differential equations of purely discontinuous Markoff processes. Trans. Amer. Math. Soc. *48* (1940), 488–515.

[7] J. Flynn: On optimality criteria for dynamic programs with long finite horizons. J. Math. Anal. Appl. *76* (1980), 202–208.

[8] M. K. Ghosh and S. I. Marcus: On strong average optimality of Markov decision processes with unbounded costs. Oper. Res. Lett. *11* (1992), 99–104.

[9] M. K. Ghosh and S. Saha: Continuous-time controlled jump Markov processes on the finite horizon. In: Optimization, Control, and Applications of Stochastic Systems (D. Hernández-Hernández and J. A. Minjárez-Sosa, eds.), Springer, New York 2012, pp. 99–109.

[10] I. I. Gihman and A. V. Skohorod: Controlled Stochastic Processes. Springer, Berlin 1979.

[11] X. P. Guo and U. Rieder: Average optimality for continuous-time Markov decision processes in Polish spaces. Ann. Appl. Probab. *16* (2006), 730–756.

[12] X. P. Guo: Continuous-time Markov decision processes with discounted rewards: the case of Polish spaces. Math. Oper. Res. *32* (2007), 73–87.

[13] X. P. Guo and O. Hernández-Lerma: Continuous-Time Markov Decision Processes: Theory and Applications. Springer, Berlin 2009.

[14] X.P. Guo and L. E. Ye: New discount and average opti mality conditions for continuous-time Markov decision processes. Adv. in Appl. Probab. *42* (2010), 953–985.

[15] O. Hernández-Lerma and J. B. Lasserre: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer, New York 1996.

[16] O. Hernández-Lerma and J. B. Lasserre: Further Topics on Discrete-Time Markov Control Processes. Springer, New York 1999.

[17] S. P. Meyn and R. L. Tweedie: Computable bounds for geometric convergence rates of Markov chains. Ann. Appl. Probab. *4* (1994), 981–1011.

[18] B. L. Miller: Finite state continuous time Markov decision processes with finite planning horizon. SIAM J. Control *6* (1968), 266–280.

[19] S. R. Pliska: Controlled jump processes. Stochastic Process. Appl. *3* (1975), 259–282.

[20] M. L. Puterman: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley, New York 1994.

[21] L. E. Ye and X. P. Guo: New sufficient conditions for average optimality in continuous-time Markov decision processes. Math. Methods Oper. Res. *72* (2010), 75–94.

[22] A. A. Yushkevich: Controlled jump Markov models. Theory Probab. Appl. *25* (1980), 244–266.

[23] Q. X. Zhu: Average optimality inequality for continuous-time Markov decision processes in Polish spaces. Math. Methods Oper. Res. *66* (2007), 299–313.

[24] Q.X. Zhu: Average optimality for continuous-time Markov decision processes with a policy iteration approach. J. Math. Anal. Appl. *339* (2008), 691–704.

Qingda Wei, School of Economics and Finance, Huaqiao University, Quanzhou, 362021. P. R. China.
e-mail: weiqd@hqu.edu.cn

Xian Chen, School of Mathematical Sciences, Peking University, Beijing, 100871. P. R. China.
e-mail: chenxian@amss.ac.cn