CrossMark

ORIGINAL ARTICLE

# Continuous-time Markov decision processes with risk-sensitive finite-horizon cost criterion

**Qingda Wei[1]**

**Abstract** This paper studies continuous-time Markov decision processes with a denumerable state space, a Borel action space, bounded cost rates and possibly unbounded transition rates under the risk-sensitive finite-horizon cost criterion. We give the suitable optimality conditions and establish the Feynman–Kac formula, via which the existence and uniqueness of the solution to the optimality equation and the existence of an optimal deterministic Markov policy are obtained. Moreover, employing a technique of the finite approximation and the optimality equation, we present an iteration method to compute approximately the optimal value and an optimal policy, and also give the corresponding error estimations. Finally, a controlled birth and death system is used to illustrate the main results.

**Keywords** Continuous-time Markov decision processes · Risk-sensitive finite-horizon cost criterion · Unbounded transition rates · Feynman–Kac formula · Finite approximation

**Mathematics Subject Classification** 93E20 · 90C40

## 1 Introduction

Continuous-time Markov decision processes (CTMDPs) have been widely studied since they have wide applications, such as the controlled queueing system, the control of the epidemic, and telecommunication; see, for instance, Puterman (1994), Kitaev

✉ Qingda Wei
weiqd@hqu.edu.cn

[1] School of Economics and Finance, Huaqiao University, Quanzhou 362021, People's Republic of China

🗘 Springer

and Rykov (1995), Guo and Hernández-Lerma (2009), Prieto-Rumeau and Lorenzo (2010), Prieto-Rumeau and Hernández-Lerma (2012), Guo and Zhang (2014) and Guo et al. (2015). The time interval in real life is always finite, so the finite-horizon optimality criterion is a commonly used criterion. The existence of optimal policies under the finite-horizon expected total cost criterion has been discussed under the different sets of optimality conditions; see, for instance, Miller (1968), Wei and Chen (2014) and Guo et al. (2015). However, the finite-horizon expected total cost criterion is risk-neutral and cannot reflect the attitude of a decision-maker to the risk. In the real-world applications, a decision-maker is usually risk-sensitive, that is, he/she is either risk-seeking or risk-averse. Thus, it is necessary for us to take the risk-sensitivity of a decision-maker into consideration in the definition of the optimality criterion. The risk-sensitive optimality criteria employ the exponential utility function to character-ize the risk-sensitivity of a decision-maker. When the risk-sensitivity coefficient of the exponential utility function takes positive (negative) values, the decision-maker is risk-averse (risk-seeking). For discrete-time MDPs, the risk-sensitive optimality criteria have been widely studied; see, for instance, Di Masi and Stettner (2007), Jaśkiewicz (2007), Cavazos-Cadena and Hernández-Hernández (2011) and the refer-ences therein. But there exist few works on the risk-sensitive optimality criteria for CTMDPs. Ghosh and Saha (2014) studied the risk-sensitive finite-horizon cost crite-rion and the risk-sensitive infinite-horizon discounted cost and average cost criteria for CTMDPs. More precisely, Ghosh and Saha (2014) also used the exponential utility function to characterize the risk-sensitivity of a decision-maker, dealt with the case of a denumerable state space, a positive risk-sensitivity coefficient, the nonnegative and bounded cost rates and the bounded transition rates, and obtained the existence of optimal policies. The positive risk-sensitivity coefficient indicates that the decision-maker is risk-averse in Ghosh and Saha (2014). However, the decision-maker may be risk-seeking in real life. Moreover, the transition rates in the controlled queueing system and the control of epidemic are usually unbounded; see, for instance, Guo and Hernández-Lerma (2009), Prieto-Rumeau and Lorenzo (2010), Prieto-Rumeau and Hernández-Lerma (2012), Guo and Zhang (2014) and Guo et al. (2015). Therefore, it is desirable for us to investigate the risk-sensitive optimality criteria for CTMDPs in which the risk-sensitivity coefficient can take positive and negative values and the transition rates are allowed to be unbounded.

In this paper we further study the risk-sensitive finite-horizon cost criterion for CTMDPs in Ghosh and Saha (2014). The state space is denumerable and the action space is a Borel space. The cost rates and the risk-sensitivity coefficient can take positive and negative values. The transition rates are allowed to be unbounded. Under the drift condition on the transition rates and the boundedness condition on the cost rates, we obtain the Feynman–Kac formula for CTMDPs which plays a crucial role in the proof of the existence of optimal policies (see Theorem 3.1). Besides the conditions for the Feynman–Kac formula, the standard continuity and compactness conditions are required for the existence of optimal policies. Under these mild conditions, we show the existence and uniqueness of the solution to the optimality equation via constructing an approximating sequence of bounded transition rates and using the Feynman–Kac formula, and then obtain the existence of an optimal deterministic Markov policy from the optimality equation (see Theorem 4.1).

On the other hand, we cannot obtain the precise forms of the optimal value and an optimal policy in general. Thus, it is of great importance for us to study the numerical methods. There are a few works dealing with the approximate computations of the risk-neutral optimality criteria for CTMDPs with the denumerable states and unbounded transition rates. van Dijk (1988, 1989) gave an approximate method for the finite-horizon expected total cost criterion via a technique of time discretization. Prieto-Rumeau and Lorenzo (2010) and Prieto-Rumeau and Hernández-Lerma (2012) studying the expected average reward criterion and the expected discounted reward criterion, respectively, proposed the numerical approximations by employing the finite truncation approach. Guo and Zhang (2014) discussed the finite approximation for the discounted CTMDPs with constraints by using a technique of occupation measures and the finite truncation method. In this paper we focus on the approximate computations of the optimal value and an optimal policy for the risk-sensitive finite-horizon cost criterion, which are not involved in Ghosh and Saha (2014). Since the state space is a denumerable set and the set of all admissible actions may be uncountable in the original control model $\mathcal{M}$, we construct a sequence of the control models $\{\mathcal{M}_n, n \geq 1\}$ with finite states and finite admissible actions to design a tractable numerical method. More specifically, applying the technique of the finite truncation in Prieto-Rumeau and Lorenzo (2010), Prieto-Rumeau and Hernández-Lerma (2012) and Guo and Zhang (2014) to the state space and the transition rates of the model $\mathcal{M}$, we give the corresponding elements of the model $\mathcal{M}_n$. Moreover, by choosing the set of all admissible actions of the model $\mathcal{M}_n$ satisfying a certain condition, we show that the optimal value of the model $\mathcal{M}_n$ converges to that of the model $\mathcal{M}$ and obtain the corresponding error estimation under the conditions which are stronger than those for the existence of optimal policies (see Theorem 5.1). Basing on this convergence and employing the optimality equation, we present an iteration method for the approximate computations of the optimal value and an optimal policy of the model $\mathcal{M}$ and give the corresponding error estimations (see Theorem 5.2). It should be mentioned that all the results on the numerical method for the risk-sensitive finite-horizon cost criterion are new. Finally, a controlled birth and death system which satisfies all the conditions in this paper is used to illustrate the applications of the risk-sensitive finite-horizon cost criterion and the numerical method.

The rest of this paper is organized as follows. In Sect. 2, we introduce the control model and the risk-sensitive finite-horizon cost criterion. In Sect. 3, we give the optimality conditions and establish the Feynman–Kac formula. In Sect. 4, we prove the existence of optimal policies via the optimality equation approach. In Sect. 5, we present an iteration method for the approximate computations of the optimal value and an optimal policy. In Sect. 6, we use a controlled birth and death system to illustrate the main results. In Sect. 7, we conclude with some remarks.

## 2 The control model

The control model under consideration is a five-tuple

$$\mathcal{M} := \{S, A, (A(i), i \in S), q(j|i, a), c(i, a)\}.$$

- The state space $S$ is the set of all nonnegative integers.
- The action space $A$ is a Borel space with the Borel $\sigma$-algebra $\mathcal{B}(A)$.
- $A(i) \in \mathcal{B}(A)$ denotes the set of all admissible actions in state $i \in S$. Let $K := \{(i, a) | i \in S, a \in A(i)\}$.
- The transition rate $q(j|i, a)$ is measurable in $a \in A(i)$ for each fixed $i, j \in S$. It satisfies $q(j|i, a) \geq 0$ for all $(i, a) \in K$ and $j \neq i$. Moreover, we assume that the transition rate is conservative and stable, i.e., $\sum_{j \in S} q(j|i, a) = 0$ for all $(i, a) \in K$, and $q^*(i) := \sup_{a \in A(i)} |q(i|i, a)| < \infty$ for all $i \in S$.
- The real-valued cost rate function $c(i, a)$ is measurable in $a \in A(i)$ for each $i \in S$.

Set $S_\infty := S \cup \{i_\infty\}$ with an isolated point $i_\infty \notin S$, $\mathbb{R}_+ := (0, +\infty)$, $\Omega^0 := (S \times \mathbb{R}_+)^\infty$ and $\Omega := \Omega^0 \cup \{(i_0, \theta_1, i_1, \ldots, \theta_{m-1}, i_{m-1}, \infty, i_\infty, \infty, i_\infty, \ldots) | i_0 \in S, \ i_l \in S, \ \theta_l \in \mathbb{R}_+$ for each $1 \leq l \leq m - 1, \ m \geq 2\}$. Thus, we obtain a measurable space $(\Omega, \mathcal{F})$ in which $\mathcal{F}$ denotes the Borel $\sigma$-algebra of $\Omega$. For each $\omega = (i_0, \theta_1, i_1, \ldots) \in \Omega$, define $X_0(\omega) := i_0, T_0(\omega) := 0, X_m(\omega) := i_m, T_m(\omega) := \theta_1 + \theta_2 + \cdots + \theta_m$ for $m \geq 1, T_\infty(\omega) := \lim_{m \to \infty} T_m(\omega)$, and the state process

$$\xi_t(\omega) := \sum_{m \geq 0} I_{\{T_m \leq t < T_{m+1}\}} i_m + I_{\{T_\infty \leq t\}} i_\infty \quad \text{for } t \geq 0,$$

where $I_D$ represents the indicator function of a set $D$. The process after $T_\infty$ is regarded to be absorbed in the state $i_\infty$. Hence, we write $q(i_\infty|i_\infty, a_\infty) = 0, c(i_\infty, a_\infty) = 0, A(i_\infty) := \{a_\infty\}, A_\infty := A \cup \{a_\infty\}$, where $a_\infty$ is an isolated point. Let $\mathcal{F}_t := \sigma(\{T_m \leq s, X_m = i\} : i \in S, s \leq t, m \geq 0)$ for $t \geq 0, \mathcal{F}_{s-} := \bigvee_{0 \leq t < s} \mathcal{F}_t$, and $\mathcal{P} := \sigma(\{D \times \{0\}, D \in \mathcal{F}_0\} \cup \{D \times (s, \infty), D \in \mathcal{F}_{s-}, s > 0\})$ which denotes the $\sigma$-algebra of predictable sets on $\Omega \times [0, \infty)$ related to $\{\mathcal{F}_t\}_{t \geq 0}$.

In order to define the optimality criterion, we introduce the definition of a policy below.

**Definition 2.1** A $\mathcal{P}$-measurable transition probability $\pi(\cdot|\omega, t)$ on $(A_\infty, \mathcal{B}(A_\infty))$, concentrated on $A(\xi_{t-}(\omega))$ is called a randomized Markov policy if there exists a kernel $\varphi$ on $A_\infty$ given $S_\infty \times [0, \infty)$ such that $\pi(\cdot|\omega, t) = \varphi(\cdot|\xi_{t-}(\omega), t)$. A policy $\pi$ is said to be deterministic Markov if there exists a measurable function $f$ on $[0, \infty) \times S_\infty$ satisfying $f(t, i) \in A(i)$ for all $(t, i) \in [0, \infty) \times S_\infty$ and $\pi(\cdot|\omega, t) = \delta_{f(t, \xi_{t-}(\omega))}(\cdot)$, where $\delta_x(\cdot)$ is the Dirac measure concentrated at the point $x$.

The class of all randomized Markov policies and the class of all deterministic Markov policies are denoted by $\Pi$ and $\Pi^D$, respectively.

For each $\pi \in \Pi$, we define the random measure

$$\nu^\pi(\omega, dt, j) := \int_A q(j|\xi_{t-}(\omega), a)\pi(da|\xi_{t-}(\omega), t) I_{\{j \neq \xi_{t-}(\omega)\}} dt$$

for all $j \in S$, which is predictable and satisfies $\nu^\pi(\omega, \{t\} \times S) = \nu^\pi(\omega, [T_\infty, \infty) \times S) = 0$. Hence, for any $\pi \in \Pi$ and any initial state $i \in S$, employing Theorem 4.27 in Kitaev and Rykov (1995), we obtain the existence of a unique probability measure $P_i^\pi$ on $(\Omega, \mathcal{F})$. Moreover, with respect to $P_i^\pi$, $\nu^\pi$ is the dual predictable projection of

the random measure on $\mathbb{R}_+ \times S$

$$\mu(\omega, dt, j) = \sum_{m \geq 1} I_{\{T_m < \infty\}} I_{\{X_m = j\}} \delta_{T_m}(dt)$$

for all $j \in S$. Let $E_i^\pi$ be the expectation operator with respect to $P_i^\pi$.

For each $\lambda \neq 0$, the exponential utility function $U_\lambda$ on $\mathbb{R} := (-\infty, \infty)$ is given by $U_\lambda(x) = sgn(\lambda)e^{\lambda x}$ for all $x \in \mathbb{R}$, where $sgn(\lambda)$ is the sign function, i.e., if $\lambda > 0$, $sgn(\lambda) = 1$; if $\lambda < 0$, $sgn(\lambda) = -1$. The constant $\lambda$ is called the risk-sensitivity coefficient. If $\lambda > 0$ $(\lambda < 0)$, the decision-maker is risk-averse (risk-seeking); see the detailed discussions in Cavazos-Cadena and Hernández-Hernández (2011).

Fix an arbitrary risk-sensitivity coefficient $\lambda \neq 0$ and the length of the horizon $T > 0$ throughout the paper. Following the ideas for the definitions of the discrete-time risk-sensitive optimality criteria in Cavazos-Cadena and Hernández-Hernández (2011), for any $\pi \in \Pi$, $i, j \in S$ and $t \in [0, T]$, the risk-sensitive $(T - t)$-horizon cost criterion $V^\pi(t, i)$ with respect to the utility function $U_\lambda$ is defined by

$$V^\pi(t, i) := \frac{1}{\lambda} \ln E_j^\pi \left[ \exp\left( \lambda \int_t^T \int_A c(\xi_s, a)\pi(da|\xi_s, s)ds \right) \bigg| \xi_t = i \right]. \quad (2.1)$$

The corresponding optimal value function is given by

$$V^*(t, i) := \inf_{\pi \in \Pi} V^\pi(t, i) \quad \text{for all} \quad (t, i) \in [0, T] \times S.$$

Note that for each $\pi \in \Pi$, $\{\xi_t, t \geq 0\}$ is a Markov jump process. Hence, $V^\pi(t, i)$ does not depend on the state $j \in S$. Moreover, since the risk-sensitivity coefficient $\lambda$ in (2.1) can take any nonzero value, the risk-sensitive finite-horizon cost criterion in this paper takes the risk-averse and risk-seeking cases into consideration. Thus, (2.1) is a generalization of the risk-sensitive finite-horizon cost criterion in Ghosh and Saha (2014) which only deals with the positive risk-sensitivity coefficient.

**Definition 2.2** A policy $\pi^* \in \Pi$ is said to be optimal if $V^{\pi^*}(0, i) = V^*(0, i)$ for all $i \in S$.

There are three main goals in this paper: (1) Give the optimality conditions and establish the optimality equation and the existence of optimal policies for CTMDPs with the unbounded transition rates. (2) Present a tractable numerical method for the approximate computations of an optimal policy and the optimal value. (3) Analyze the accuracy of the numerical method and obtain the corresponding error estimations.

## 3 Preliminaries

In this section, we give the optimality conditions for the existence of the optimal policies and obtain the Feynman–Kac formula which is very useful in proving the main results.

To avoid the explosion of the state process $\{\xi_t, t \geq 0\}$ and ensure the finiteness of the value function $V^*$, we introduce the following assumption consisting of the drift condition on the transition rates and the boundedness condition on the cost rates; see, for instance, Guo and Hernández-Lerma (2009), Prieto-Rumeau and Lorenzo (2010), Prieto-Rumeau and Hernández-Lerma (2012), Wei and Chen (2014), Guo and Zhang (2014) and Guo et al. (2015).

**Assumption 3.1** There exist a function $w \geq 1$ on $S$ and constants $\rho_1 > 0$, $d_1 \geq 0$, $R > 0$ and $M > 0$ such that

(i) $\sum_{j \in S} w(j)q(j|i,a) \leq \rho_1 w(i) + d_1$ for all $(i,a) \in K$;
(ii) $q^*(i) \leq Rw(i)$ for all $i \in S$;
(iii) $|c(i,a)| \leq M$ for all $(i,a) \in K$.

To ensure the existence of the optimal policies, we also need the following assumption, which is called the standard continuity and compactness conditions; see, for instance, Puterman (1994), Kitaev and Rykov (1995), Guo and Hernández-Lerma (2009), Prieto-Rumeau and Lorenzo (2010), Prieto-Rumeau and Hernández-Lerma (2012), Wei and Chen (2014), Ghosh and Saha (2014), Guo and Zhang (2014) and Guo et al. (2015).

**Assumption 3.2** (i) For each $i \in S$, the set $A(i)$ is compact.
(ii) For each $i, j \in S$, the functions $c(i,a)$ and $q(j|i,a)$ are continuous in $a \in A(i)$.

Let $w$ be as in Assumption 3.1. A real-valued function $u$ on $[0, T] \times S$ is called $w$-bounded if it satisfies the norm $\|u\|_w := \sup_{(t,i) \in [0,T] \times S} \frac{|u(t,i)|}{w(i)} < \infty$. Denote by $B_w([0, T] \times S)$ the set of all $w$-bounded measurable functions on $[0, T] \times S$ and by $B([0, T] \times S)$ the set of all bounded measurable functions $u$ on $[0, T] \times S$ with the norm $\|u\| := \sup_{(t,i) \in [0,T] \times S} |u(t,i)| < \infty$. Let $\mathcal{L}([0, T] \times S) := \{u \in B([0, T] \times S) :$ for each $i \in S$, $u(\cdot, i)$ is differentiable on $[0, T]$ and $\frac{\partial u}{\partial t} \in B_w([0, T] \times S)\}$, where $\frac{\partial u}{\partial t}$ denotes the derivative of $u$ with respect to the variable $t$.

Below we give the Feynman–Kac formula for CTMDPs with the unbounded transition rates, which plays a crucial role in proving the existence of optimal policies.

**Theorem 3.1** *Suppose that Assumption* 3.1 *is satisfied. Then for each* $u \in \mathcal{L}([0, T] \times S)$, $i, j \in S$, $\pi \in \Pi$ *and* $0 \leq s \leq t \leq T$, *we have*

$$
E_j^\pi \left[ \exp \left( \lambda \int_s^t \int_A c(\xi_v, a)\pi(da|\xi_v, v)dv \right) u(t, \xi_t) \Big| \xi_s = i \right] - u(s, i)
$$
$$
= E_j^\pi \left[ \int_s^t \exp \left( \lambda \int_s^r \int_A c(\xi_v, a)\pi(da|\xi_v, v)dv \right) \left( \lambda \int_A c(\xi_r, a)\pi(da|\xi_r, r)u(r, \xi_r) \right. \right.
$$
$$
\left. \left. + \frac{\partial u}{\partial r}(r, \xi_r) + \sum_{k \in S} u(r, k) \int_A q(k|\xi_r, a)\pi(da|\xi_r, r) \right) dr \Big| \xi_s = i \right].
$$

*Proof* Fix any $i, j \in S$, $\pi \in \Pi$ and $0 \leq s \leq t \leq T$. Let $L :=$ $\sup_{(v,k)\in[0,T]\times S} |u(v,k)|$. Then by Assumption 3.1 and Lemma 6.3 in Guo and Hernández-Lerma (2009), we obtain

$$E_j^\pi \left[ \exp\left( \lambda \int_s^t \int_A c(\xi_v, a)\pi(da|\xi_v, v)dv \right) |u(t, \xi_t)| \Big| \xi_s = i \right] \leq L e^{|\lambda|MT} \text{ and}$$

$$E_j^\pi \left[ \int_s^t \exp\left( \lambda \int_s^r \int_A c(\xi_v, a)\pi(da|\xi_v, v)dv \right) \left( \left| \lambda \int_A c(\xi_r, a)\pi(da|\xi_r, r)u(r, \xi_r) \right| \right. \right.$$

$$\left. \left. + \left| \frac{\partial u}{\partial r}(r, \xi_r) \right| + \sum_{k\in S} |u(r, k)| \int_A |q(k|\xi_r, a)|\pi(da|\xi_r, r) \right) dr \Big| \xi_s = i \right]$$

$$\leq e^{|\lambda|MT} \left[ |\lambda|ML + \left\| \frac{\partial u}{\partial t} \right\|_w + 2LR \right] E_j^\pi \left[ \int_s^t w(\xi_r)dr \Big| \xi_s = i \right]$$

$$\leq e^{|\lambda|MT} \left[ |\lambda|ML + \left\| \frac{\partial u}{\partial t} \right\|_w + 2LR \right] \int_s^t \left[ e^{\rho_1(r-s)}w(i) + \frac{d_1}{\rho_1}\left( e^{\rho_1(r-s)} - 1 \right) \right] dr$$

$$\leq T e^{|\lambda|MT} \left[ |\lambda|ML + \left\| \frac{\partial u}{\partial t} \right\|_w + 2LR \right] \left( w(i) + \frac{d_1}{\rho_1} \right) e^{\rho_1 T} < \infty.$$

Since $u$ belongs to $\mathcal{L}([0, T] \times S)$, the differential mean value theorem yields that for any $0 \leq t_1 < t_2 \leq T$, there exists a constant $\widetilde{t} \in (t_1, t_2)$ such that

$$|u(t_2, i) - u(t_1, i)| = (t_2 - t_1)\left| \frac{\partial u}{\partial t}(\widetilde{t}, i) \right| \leq (t_2 - t_1)\left\| \frac{\partial u}{\partial t} \right\|_w w(i). \tag{3.1}$$

Similarly, for any $0 \leq t_2 < t_1 \leq T$, we can get

$$|u(t_2, i) - u(t_1, i)| \leq (t_1 - t_2)\left\| \frac{\partial u}{\partial t} \right\|_w w(i). \tag{3.2}$$

Thus, combining (3.1) and (3.2), we obtain

$$|u(t_1, i) - u(t_2, i)| \leq \left\| \frac{\partial u}{\partial t} \right\|_w w(i)|t_1 - t_2|$$

for all $t_1, t_2 \in [0, T]$, i.e., $u(\cdot, i)$ is Lipschitz continuous on $[0, T]$. Hence, it follows from the result in Royden (1988, p. 112) that $u(\cdot, i)$ is absolutely continuous on $[0, T]$. Moreover, we have

$$E_j^\pi \left[ \int_s^T \sum_{k\in S} |u(r, k) - u(r, \xi_{r-})| \mu(dr, k) \Big| \xi_s = i \right]$$

$$\leq 2L E_j^\pi [\mu((s, T] \times S)|\xi_s = i]$$

$$= 2L E_j^\pi [\nu^\pi((s, T] \times S)|\xi_s = i]$$

$$= 2L E_j^\pi \left[ \int_s^T \int_A |q(\xi_r|\xi_r, a)|\pi(da|\xi_r, r)dr \Big| \xi_s = i \right]$$

$$\leq 2LRE_j^{\pi}\left[\int_s^T w(\xi_r)dr\,\middle|\,\xi_s = i\right]$$

$$\leq 2LRT\left(w(i) + \frac{d_1}{\rho_1}\right)e^{\rho_1 T},$$

where the first equality is due to the fact that $\nu^{\pi}$ is the dual predictable projection of $\mu$, and the last inequality follows from Lemma 6.3 in Guo and Hernández-Lerma (2009). Thus, direct calculations give

$$E_j^{\pi}\left[\int_s^t \lambda \exp\left(\lambda\int_s^r\int_A c(\xi_v,a)\pi(da|\xi_v,v)dv\right)\right.$$

$$\left.\int_A c(\xi_r,a)\pi(da|\xi_r,r)u(r,\xi_r)dr\,\middle|\,\xi_s = i\right]$$

$$= E_j^{\pi}\left[\exp\left(\lambda\int_s^t\int_A c(\xi_v,a)\pi(da|\xi_v,v)dv\right)u(t,\xi_t)\,\middle|\,\xi_s = i\right] - u(s,i)$$

$$- E_j^{\pi}\left[\int_s^t \exp\left(\lambda\int_s^r\int_A c(\xi_v,a)\pi(da|\xi_v,v)dv\right)du(r,\xi_r)\,\middle|\,\xi_s = i\right]$$

$$= E_j^{\pi}\left[\exp\left(\lambda\int_s^t\int_A c(\xi_v,a)\pi(da|\xi_v,v)dv\right)u(t,\xi_t)\,\middle|\,\xi_s = i\right] - u(s,i)$$

$$- E_j^{\pi}\left[\int_s^t \exp\left(\lambda\int_s^r\int_A c(\xi_v,a)\pi(da|\xi_v,v)dv\right)\right.$$

$$\left.\times\left(\frac{\partial u}{\partial r}(r,\xi_r) + \sum_{k\in S}u(r,k)\int_A q(k|\xi_r,a)\pi(da|\xi_r,r)\right)dr\,\middle|\,\xi_s = i\right]$$

$$- E_j^{\pi}\left[\int_s^t \exp\left(\lambda\int_s^r\int_A c(\xi_v,a)\pi(da|\xi_v,v)dv\right)\right.$$

$$\left.\times\sum_{k\in S}\big(u(r,k) - u(r,\xi_{r-})\big)\big(\mu(dr,k) - \nu^{\pi}(dr,k)\big)\,\middle|\,\xi_s = i\right]$$

$$= E_j^{\pi}\left[\exp\left(\lambda\int_s^t\int_A c(\xi_v,a)\pi(da|\xi_v,v)dv\right)u(t,\xi_t)\,\middle|\,\xi_s = i\right] - u(s,i)$$

$$- E_j^{\pi}\left[\int_s^t \exp\left(\lambda\int_s^r\int_A c(\xi_v,a)\pi(da|\xi_v,v)dv\right)\right.$$

$$\left.\times\left(\frac{\partial u}{\partial r}(r,\xi_r) + \sum_{k\in S}u(r,k)\int_A q(k|\xi_r,a)\pi(da|\xi_r,r)\right)dr\,\middle|\,\xi_s = i\right], \quad (3.3)$$

where the first equality is due to the integration by parts, the second one follows from the equation (2.9) in Confortola and Fuhrman (2014), and the third one is due to the fact that $\nu^{\pi}$ is the dual predictable projection of $\mu$. Hence, the desired assertion follows from (3.3). □

## 4 The existence of optimal policies

In this section, we establish the optimality equation and the existence of optimal policies for the risk-sensitive finite-horizon cost criterion with the unbounded transition rates.

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

(a) $|V^\pi(t, i)| \leq MT$ *and* $|V^*(t, i)| \leq MT$ *for all* $\pi \in \Pi$ *and* $(t, i) \in [0, T] \times S$.

(b) *There exists a unique solution in* $\mathcal{L}([0, T] \times S)$ *to the following equation:*

$$\begin{cases} -\frac{\partial u}{\partial t}(t, i) = sgn(\lambda) \inf_{a \in A(i)} sgn(\lambda) \left\{ \lambda c(i, a) u(t, i) + \sum_{j \in S} u(t, j) q(j|i, a) \right\}, \\ u(T, i) = 1, \end{cases}$$

(4.1)

*for all* $(t, i) \in [0, T] \times S$. *Moreover, we have* $u(t, i) = e^{\lambda V^*(t,i)}$ *for all* $(t, i) \in [0, T] \times S$.

(c) *There exists* $f^* \in \Pi^D$ *satisfying*

$$\lambda e^{\lambda V^*(t,i)} \frac{\partial V^*}{\partial t}(t, i) + \lambda c(i, f^*(t, i)) e^{\lambda V^*(t,i)} + \sum_{j \in S} e^{\lambda V^*(t,j)} q(j|i, f^*(t, i)) = 0$$

(4.2)

*for all* $(t, i) \in [0, T] \times S$ *and* $f^*$ *is an optimal deterministic Markov policy.*

*Proof* (a) By Assumption 3.1(iii), we have

$$e^{-|\lambda|MT} \leq E_j^\pi \left[ \exp\left( \lambda \int_t^T \int_A c(\xi_s, a) \pi(da|\xi_s, s) ds \right) \Big| \xi_t = i \right] \leq e^{|\lambda|MT}$$

for all $i, j \in S$, $\pi \in \Pi$ and $t \in [0, T]$, which gives the desired assertion.

(b) We only prove the case $\lambda < 0$ because the arguments for the case $\lambda > 0$ are similar. For each integer $m \geq 1$, let $S_m := \{k \in S : w(k) \leq m\}$ and we have $S_m \uparrow S$. For each $(i, a) \in K$, define

$$q^{(m)}(j|i, a) := \begin{cases} q(j|i, a), & \text{if } i \in S_m, \\ 0, & \text{if } i \notin S_m, \end{cases}$$

(4.3)

and an operator $\Gamma_m$ on $B([0, T] \times S)$ as follows:

$$\Gamma_m g(t, i) := e^{\alpha t} + e^{\alpha t} \int_t^T e^{-\alpha s} \sup_{a \in A(i)} \left\{ \lambda c(i, a) g(s, i) + \sum_{j \in S} g(s, j) q^{(m)}(j|i, a) \right\} ds$$

for all $g \in B([0, T] \times S)$, where $\alpha$ is an arbitrary positive constant. Then using Assumption 3.1, we obtain

$$|\Gamma_m g(t, i)| \leq e^{\alpha T} + T e^{\alpha T} (|\lambda| M + 2Rm) \|g\|$$

for all $(t, i) \in [0, T] \times S$, which implies that $\Gamma_m$ is a map from $B([0, T] \times S)$ into itself. On the other hand, for any $g_1, g_2 \in B([0, T] \times S)$, we have

$$
\begin{aligned}
&|\Gamma_m g_1(t, i) - \Gamma_m g_2(t, i)| \\
&\leq e^{\alpha t} \int_t^T e^{-\alpha s} \sup_{a \in A(i)} \bigg| \lambda c(i, a)\big(g_1(s, i) - g_2(s, i)\big) \\
&\quad + \sum_{j \in S} \big(g_1(s, j) - g_2(s, j)\big) q^{(m)}(j|i, a) \bigg| ds \\
&\leq \frac{(|\lambda| M + 2Rm)}{\alpha} \|g_1 - g_2\|
\end{aligned}
$$

for all $(t, i) \in [0, T] \times S$. Thus, taking $\alpha = |\lambda| M + 2Rm + 1$, we see that $\Gamma_m$ is a contraction operator on $B([0, T] \times S)$. Hence, it follows from the Banach fixed point theorem that there exists $g^{(m)} \in B([0, T] \times S)$ satisfying

$$
g^{(m)}(t, i) = e^{\alpha t} + e^{\alpha t} \int_t^T e^{-\alpha s} \sup_{a \in A(i)} \left\{ \lambda c(i, a) g^{(m)}(s, i) + \sum_{j \in S} g^{(m)}(s, j) q^{(m)}(j|i, a) \right\} ds
$$

for all $(t, i) \in [0, T] \times S$ and $\alpha = |\lambda| M + 2Rm + 1$. Let $u^{(m)}(t, i) := e^{-(|\lambda| M + 2Rm + 1)t} g^{(m)}(t, i)$. The last equality can be rewritten as

$$
u^{(m)}(t, i) = 1 + \int_t^T \sup_{a \in A(i)} \left\{ \lambda c(i, a) u^{(m)}(s, i) + \sum_{j \in S} u^{(m)}(s, j) q^{(m)}(j|i, a) \right\} ds.
$$
(4.4)

Note that for each $i \in S$, $u^{(m)}(\cdot, i)$ is absolutely continuous on $[0, T]$ and $u^{(m)}(T, i) = 1$. Moreover, it follows from Assumption 3.2 that for each $i \in S$,

$$
\sup_{a \in A(i)} \left\{ \lambda c(i, a) u^{(m)}(s, i) + \sum_{j \in S} u^{(m)}(s, j) q^{(m)}(j|i, a) \right\}
$$

is continuous in $s \in [0, T]$. Thus, the equality (4.4) gives

$$
-\frac{\partial u^{(m)}}{\partial t}(t, i) = \sup_{a \in A(i)} \left\{ \lambda c(i, a) u^{(m)}(t, i) + \sum_{j \in S} u^{(m)}(t, j) q^{(m)}(j|i, a) \right\}
$$
(4.5)

and $\left| \frac{\partial u^{(m)}}{\partial t}(t, i) \right| \leq (|\lambda| M + 2R) \|g^{(m)}\| w(i)$ for all $(t, i) \in [0, T] \times S$. Hence, we have $u^{(m)} \in \mathcal{L}([0, T] \times S)$ for all $m \geq 1$. For each $i \in S$, $\pi \in \Pi$ and $m \geq 1$, denote by $P_i^{\pi, m}$ the probability measure corresponding to $q^{(m)}(j|i, a)$ and by $E_i^{\pi, m}$ the expectation operator with respect to $P_i^{\pi, m}$. On one hand, Assumption 3.2 and the

measurable selection theorem in Hernández-Lerma and Lasserre (1999, p. 50) give the existence of a measurable function $f_m$ on $[0, T] \times S$ satisfying

$$-\frac{\partial u^{(m)}}{\partial t}(t, i) = \lambda c(i, f_m(t, i))u^{(m)}(t, i) + \sum_{j \in S} u^{(m)}(t, j)q^{(m)}(j|i, f_m(t, i))$$

for all $(t, i) \in [0, T] \times S$. Then using the last equality, we obtain

$$-E_j^{f_m, m}\left[\int_t^T \exp\left(\lambda \int_t^s c(\xi_v, f_m(v, \xi_v))dv\right)\frac{\partial u^{(m)}}{\partial s}(s, \xi_s)ds \,\middle|\, \xi_t = i\right]$$

$$= E_j^{f_m, m}\left[\int_t^T \exp\left(\lambda \int_t^s c(\xi_v, f_m(v, \xi_v))dv\right)\left(\lambda c(\xi_s, f_m(s, \xi_s))u^{(m)}(s, \xi_s)\right.\right.$$

$$\left.\left. + \sum_{k \in S} u^{(m)}(s, k)q^{(m)}(k|\xi_s, f_m(s, \xi_s))\right)ds \,\middle|\, \xi_t = i\right],$$

which together with Theorem 3.1 and $u^{(m)}(T, i) = 1$ yields

$$u^{(m)}(t, i) = E_j^{f_m, m}\left[\exp\left(\lambda \int_t^T c(\xi_s, f_m(s, \xi_s))ds\right)\,\middle|\, \xi_t = i\right] \tag{4.6}$$

for all $i, j \in S$ and $t \in [0, T]$. On the other hand, for any $\pi \in \Pi$, by (4.5) we have

$$-\frac{\partial u^{(m)}}{\partial s}(s, \xi_s) \geq \lambda \int_A c(\xi_s, a)\pi(da|\xi_s, s)u^{(m)}(s, \xi_s)$$

$$+ \sum_{j \in S} u^{(m)}(s, j)\int_A q^{(m)}(j|\xi_s, a)\pi(da|\xi_s, s)$$

for all $s \in [0, T]$. Then employing the last inequality and following the similar arguments of (4.6), we obtain

$$u^{(m)}(t, i) \geq E_j^{\pi, m}\left[\exp\left(\lambda \int_t^T \int_A c(\xi_s, a)\pi(da|\xi_s, s)ds\right)\,\middle|\, \xi_t = i\right] \tag{4.7}$$

for all $i, j \in S$, $t \in [0, T]$ and $\pi \in \Pi$. Thus, combining (4.6) and (4.7), we get

$$u^{(m)}(t, i) = \sup_{\pi \in \Pi} E_j^{\pi, m}\left[\exp\left(\lambda \int_t^T \int_A c(\xi_s, a)\pi(da|\xi_s, s)ds\right)\,\middle|\, \xi_t = i\right],$$

which implies

$$0 < u^{(m)}(t, i) \leq e^{|\lambda|MT} \tag{4.8}$$

for all $i, j \in S$ and $t \in [0, T]$. Moreover, observe that Assumption 3.1(i) and (ii) still hold with $q^{(m)}(j|i, a)$ in lieu of $q(j|i, a)$. By Assumption 3.1 and (4.8), we have

$$\left| \sup_{a \in A(i)} \left\{ \lambda c(i, a) u^{(m)}(t, i) + \sum_{j \in S} u^{(m)}(t, j) q^{(m)}(j|i, a) \right\} \right| \leq (|\lambda| M + 2Rw(i)) e^{|\lambda| MT}$$

(4.9)

for all $(t, i) \in [0, T] \times S$. Let $M_i := (|\lambda| M + 2Rw(i)) e^{|\lambda| MT}$. Then for each $i \in S$ and any $\varepsilon > 0$, there exists $\eta := \frac{\varepsilon}{M_i} > 0$ such that

$$|u^{(m)}(t_1, i) - u^{(m)}(t_2, i)|$$
$$\leq \int_{\min\{t_1, t_2\}}^{\max\{t_1, t_2\}} \left| \sup_{a \in A(i)} \left\{ \lambda c(i, a) u^{(m)}(s, i) + \sum_{j \in S} u^{(m)}(s, j) q^{(m)}(j|i, a) \right\} \right| ds$$
$$\leq M_i |t_1 - t_2| \leq \varepsilon$$

for all $m \geq 1$, $t_1, t_2 \in [0, T]$ and $|t_1 - t_2| \leq \eta$, where the first and second inequalities follow from (4.4) and (4.9), respectively. Thus, for each $i \in S$, $\{u^{(m)}(\cdot, i), m \geq 1\}$ is uniformly bounded and equicontinuous. Hence, employing the Ascoli–Arzela theorem in Royden (1988, p. 169) and the denumerability of $S$, we obtain the existence of a subsequence of $\{m\}$ (still denoted by $\{m\}$) such that $\lim_{m \to \infty} u^{(m)}(t, i) =: u(t, i)$, $\|u\| \leq e^{|\lambda| MT}$ and $u(T, i) = 1$ for all $(t, i) \in [0, T] \times S$. Below we will show that for each $(t, i) \in [0, T] \times S$,

$$\lim_{m \to \infty} \sup_{a \in A(i)} \left\{ \lambda c(i, a) u^{(m)}(t, i) + \sum_{j \in S} u^{(m)}(t, j) q^{(m)}(j|i, a) \right\}$$
$$= \sup_{a \in A(i)} \left\{ \lambda c(i, a) u(t, i) + \sum_{j \in S} u(t, j) q(j|i, a) \right\}.$$

(4.10)

For each $(t, i) \in [0, T] \times S$ and $m \geq 1$, Assumption 3.2 implies that there exists $a_m^{t,i} \in A(i)$ satisfying

$$\sup_{a \in A(i)} \left| \lambda c(i, a) u^{(m)}(t, i) + \sum_{j \in S} u^{(m)}(t, j) q^{(m)}(j|i, a) - \lambda c(i, a) u(t, i) \right.$$
$$\left. - \sum_{j \in S} u(t, j) q(j|i, a) \right| = \left| \lambda c(i, a_m^{t,i}) u^{(m)}(t, i) + \sum_{j \in S} u^{(m)}(t, j) q^{(m)}(j|i, a_m^{t,i}) \right.$$
$$\left. - \lambda c(i, a_m^{t,i}) u(t, i) - \sum_{j \in S} u(t, j) q(j|i, a_m^{t,i}) \right|.$$

(4.11)

Then we have

$$
\left| \lambda c(i, a_m^{t,i}) u^{(m)}(t, i) + \sum_{j \in S} u^{(m)}(t, j) q^{(m)}(j|i, a_m^{t,i}) - \lambda c(i, a_m^{t,i}) u(t, i) \right.
$$

$$
\left. - \sum_{j \in S} u(t, j) q(j|i, a_m^{t,i}) \right| \to 0 \tag{4.12}
$$

as $m \to \infty$. In fact, if (4.12) does not hold, there exist $\varepsilon_0 > 0$ and a subsequence $\{m_l\}$ of $\{m\}$ such that

$$
\left| \lambda c(i, a_{m_l}^{t,i}) u^{(m_l)}(t, i) + \sum_{j \in S} u^{(m_l)}(t, j) q^{(m_l)}(j|i, a_{m_l}^{t,i}) - \lambda c(i, a_{m_l}^{t,i}) u(t, i) \right.
$$

$$
\left. - \sum_{j \in S} u(t, j) q(j|i, a_{m_l}^{t,i}) \right| \geq \varepsilon_0. \tag{4.13}
$$

The compactness of $A(i)$ gives the existence of a subsequence of $\{m_l\}$ (still denoted by $\{m_l\}$) such that $a_{m_l}^{t,i}$ converges to some $a^{t,i} \in A(i)$. Thus, it follows from Assumption 3.2(ii) and (4.3) that

$$
\left| \lambda c(i, a_{m_l}^{t,i}) u^{(m_l)}(t, i) + \sum_{j \in S} u^{(m_l)}(t, j) q^{(m_l)}(j|i, a_{m_l}^{t,i}) - \lambda c(i, a_{m_l}^{t,i}) u(t, i) \right.
$$

$$
\left. - \sum_{j \in S} u(t, j) q(j|i, a_{m_l}^{t,i}) \right| \to 0
$$

as $l \to \infty$, which yields a contradiction to (4.13). Hence, (4.12) is true. Therefore, (4.11), (4.12) and the inequality

$$
\left| \sup_{a \in A(i)} \left\{ \lambda c(i, a) u^{(m)}(t, i) + \sum_{j \in S} u^{(m)}(t, j) q^{(m)}(j|i, a) \right\} \right.
$$

$$
\left. - \sup_{a \in A(i)} \left\{ \lambda c(i, a) u(t, i) + \sum_{j \in S} u(t, j) q(j|i, a) \right\} \right|
$$

$$
\leq \sup_{a \in A(i)} \left| \lambda c(i, a) u^{(m)}(t, i) + \sum_{j \in S} u^{(m)}(t, j) q^{(m)}(j|i, a) - \lambda c(i, a) u(t, i) \right.
$$

$$
\left. - \sum_{j \in S} u(t, j) q(j|i, a) \right|
$$

for all $(t, i) \in [0, T] \times S$, imply (4.10). Moreover, employing (4.4), (4.9), (4.10) and the dominated convergence theorem, we get

$$u(t, i) = 1 + \int_t^T \sup_{a \in A(i)} \left\{ \lambda c(i, a) u(s, i) + \sum_{j \in S} u(s, j) q(j|i, a) \right\} ds,$$

which together with Assumption 3.2 yields that the derivative of $u(t, i)$ with respect to the variable $t$ exists for all $(t, i) \in [0, T] \times S$. Thus, (4.1) follows from the last equality. Note that Assumption 3.1, $\|u\| \leq e^{|\lambda| M T}$ and (4.1) imply

$$\left| \frac{\partial u}{\partial t}(t, i) \right| \leq (|\lambda| M + 2R) e^{|\lambda| M T} w(i)$$

for all $(t, i) \in [0, T] \times S$. Hence, we obtain $u \in \mathcal{L}([0, T] \times S)$. Using (4.1) and following the similar arguments of (4.6) and (4.7), there exists $f^* \in \Pi^D$ satisfying

$$\frac{\partial u}{\partial t}(t, i) + \lambda c(i, f^*(t, i)) u(t, i) + \sum_{j \in S} u(t, j) q(j|i, f^*(t, i)) = 0 \qquad (4.14)$$

and

$$u(t, i) = e^{\lambda V^{f^*}(t, i)} \geq E_j^\pi \left[ \exp \left( \lambda \int_t^T \int_A c(\xi_s, a) \pi(da|\xi_s, s) ds \right) \Big| \xi_t = i \right] \qquad (4.15)$$

for all $i, j \in S, t \in [0, T]$ and $\pi \in \Pi$. Thus, it follows from (4.14) and (4.15) that $u(t, i) = e^{\lambda V^*(t, i)} = e^{\lambda V^{f^*}(t, i)}$ for all $(t, i) \in [0, T] \times S$, and $f^*$ is optimal and satisfies (4.2). Finally, the uniqueness of the solution to (4.1) follows from the similar arguments of (4.6) and (4.7).

(c) The assertion follows directly from the proof of part (b). □

*Remark 4.1* The optimality equation of the risk-sensitive finite-horizon cost criterion has been established in Ghosh and Saha (2014) for the case of a positive risk-sensitivity coefficient, nonnegative and bounded cost rates and bounded transition rates. Theorem 4.1 extends the results in Ghosh and Saha (2014) to the case in which the risk-sensitivity coefficient and the cost rates can take positive and negative values, and the transition rates are allowed to be unbounded.

## 5 Finite approximation

In this section, we use the optimality equation established in Theorem 4.1 to give a finite approximation method for the approximate computations of an optimal policy and the optimal value. In order to obtain the error estimations of the approximate computations, we introduce the following conditions which are stronger than those in Sect. 3.

**Assumption 5.1** (i) The function $w$ in Assumption 3.1 is nondecreasing and satisfies $\lim_{i \to \infty} w(i) = \infty$.

(ii) There exist constants $\rho_2 > 0$ and $d_2 \geq 0$ such that

$$\sum_{j \in S} w^2(j) q(j|i, a) \leq \rho_2 w^2(i) + d_2 \quad \text{for all} \quad (i, a) \in K.$$

(iii) For any $i, j \in S$, there exist constants $L_i > 0$ and $L_{ij} > 0$ such that

$$|c(i, a) - c(i, b)| \leq L_i d_A(a, b) \quad \text{and} \quad |q(j|i, a) - q(j|i, b)| \leq L_{ij} d_A(a, b)$$

for all $a, b \in A(i)$, where $d_A$ represents the metric of the space $A$.

For each integer $n \geq 1$, we define the control model

$$\mathcal{M}_n := \{S_n, A, (A_n(i), i \in S_n), q_n(j|i, a), c(i, a)\}.$$

- The state space is given by $S_n := \{0, 1, \ldots, j_n\}$ and the increasing sequence $\{j_n, n \geq 1\}$ satisfies $\lim_{n \to \infty} j_n = \infty$.
- The action space is given by $A$ as in the model $\mathcal{M}$.
- The set of all admissible actions in the state $i \in S_n$ is given by an arbitrary finite set $A_n(i)$.
- For each $(i, a) \in K_n := \{(i, a)|i \in S_n, a \in A_n(i)\}$ and $j \in S_n$, the transition rate $q_n(j|i, a)$ is given by

$$q_n(j|i, a) := \begin{cases} q(j|i, a), & \text{if } j \neq j_n, \\ \sum_{k \geq j_n} q(k|i, a), & \text{if } j = j_n. \end{cases} \tag{5.1}$$

- The cost rate function is given by the restriction of $c$ in the model $\mathcal{M}$ to $K_n$.

Let $\Pi_n$ and $\Pi_n^D$ be the set of all randomized Markov policies and the set of all deterministic Markov policies for $\mathcal{M}_n$, respectively. For each $i \in S_n$ and $\pi \in \Pi_n$, Theorem 4.27 in Kitaev and Rykov (1995) gives the existence of a probability measure $P_n^{i,\pi}$ associated with the model $\mathcal{M}_n$. Let $E_n^{i,\pi}$ be the expectation operator with respect to $P_n^{i,\pi}$. Moreover, replacing $E_i^\pi$ and $\Pi$ by $E_n^{i,\pi}$ and $\Pi_n$ in the definitions of $V^\pi$ and $V^*$, we can define the functions $V_n^\pi$ and $V_n^*$ on $[0, T] \times S_n$. Denote by $\mathcal{C}$ the set of all closed subsets of $A$. Recall that the Hausdorff metric on $\mathcal{C}$ is defined by

$$d_H(B_1, B_2) := \max \left\{ \sup_{a \in B_1} \inf_{b \in B_2} d_A(a, b), \sup_{b \in B_2} \inf_{a \in B_1} d_A(a, b) \right\}$$

for all $B_1, B_2 \in \mathcal{C}$.

Then we have the following statement about the error estimation between $V^*$ and $V_n^*$.

**Theorem 5.1** *Suppose that Assumptions 3.1, 3.2(i) and 5.1 are satisfied. If there exists a constant $\widetilde{M} > 0$ such that for each $n \geq 1$ and $i \in S_n$,*

$$d_H(A(i), A_n(i)) \leq \frac{\widetilde{M} w^2(i)}{e^{|\lambda|MT} \left( |\lambda| L_i + 2 \sum_{j=0}^{j_n-1} L_{ij} \right) w(j_n)}, \tag{5.2}$$

*then we have*

$$|V_n^*(t, i) - V^*(t, i)| \leq \frac{R_1 w^2(i)}{w(j_n)}$$

*for all $t \in [0, T]$, where $R_1 = \frac{T e^{2|\lambda|MT}}{|\lambda|} \left[ \widetilde{M} + e^{|\lambda|MT} (\rho_1 + d_1 + R) \right] \times \left[ e^{\rho_2 T} + \frac{d_2}{\rho_2} e^{\rho_2 T} \right]$.*

*Proof* Fix any $n \geq 1$, $i \in S_n$ and $t \in [0, T]$. Then it follows from Theorem 4.1 that there exists $f^* \in \Pi^D$ satisfying

$$-\lambda e^{\lambda V^*(t,i)} \frac{\partial V^*}{\partial t}(t, i) = \lambda c(i, f^*(t, i)) e^{\lambda V^*(t,i)} + \sum_{j \in S} e^{\lambda V^*(t,j)} q(j|i, f^*(t, i))$$

$$= \lambda c(i, f^*(t, i)) e^{\lambda V^*(t,i)}$$

$$+ \sum_{j=0}^{j_n-1} \left( e^{\lambda V^*(t,j)} - e^{\lambda V^*(t,j_n)} \right) q(j|i, f^*(t, i))$$

$$+ \sum_{j > j_n} \left( e^{\lambda V^*(t,j)} - e^{\lambda V^*(t,j_n)} \right) q(j|i, f^*(t, i)). \tag{5.3}$$

By Theorem 4.1, the monotonicity of $w$ and Assumption 3.1, we obtain

$$\sum_{j > j_n} \left( e^{\lambda V^*(t,j)} - e^{\lambda V^*(t,j_n)} \right) q(j|i, f^*(t, i))$$

$$\geq -e^{|\lambda|MT} \sum_{j > j_n} q(j|i, f^*(t, i))$$

$$\geq -\frac{e^{|\lambda|MT}}{w(j_n)} \sum_{j > j_n} w(j) q(j|i, f^*(t, i))$$

$$\geq -\frac{e^{|\lambda|MT}}{w(j_n)} \left[ \sum_{j \in S} w(j) q(j|i, f^*(t, i)) - w(i) q(i|i, f^*(t, i)) \right]$$

$$\geq -e^{|\lambda|MT} (\rho_1 + d_1 + R) \frac{w^2(i)}{w(j_n)}. \tag{5.4}$$

Moreover, for each $s \in [0, T]$, there exists $\widetilde{f}(s, i) \in A_n(i)$ such that

$$d_A(f^*(s, i), \widetilde{f}(s, i)) = \min_{a \in A_n(i)} d_A(f^*(s, i), a) \leq d_H(A(i), A_n(i)).$$

Thus, employing the last inequality, Assumption 5.1 and Theorem 4.1, we have

$$\lambda c(i, f^*(t,i))e^{\lambda V^*(t,i)} - \lambda c(i, \widetilde{f}(t,i))e^{\lambda V^*(t,i)} \geq - |\lambda|e^{|\lambda|MT} L_i d_A(f^*(t,i), \widetilde{f}(t,i))$$
$$\geq - |\lambda|e^{|\lambda|MT} L_i d_H(A(i), A_n(i))$$

and

$$\sum_{j=0}^{j_n-1} \left( e^{\lambda V^*(t,j)} - e^{\lambda V^*(t,j_n)} \right) \left( q(j|i, f^*(t,i)) - q(j|i, \widetilde{f}(t,i)) \right)$$

$$\geq -2e^{|\lambda|MT} \sum_{j=0}^{j_n-1} \left| q(j|i, f^*(t,i)) - q(j|i, \widetilde{f}(t,i)) \right|$$

$$\geq -2e^{|\lambda|MT} d_A(f^*(t,i), \widetilde{f}(t,i)) \sum_{j=0}^{j_n-1} L_{ij}$$

$$\geq -2e^{|\lambda|MT} d_H(A(i), A_n(i)) \sum_{j=0}^{j_n-1} L_{ij},$$

which together with (5.1)–(5.4) give

$$- \lambda e^{\lambda V^*(t,i)} \frac{\partial V^*}{\partial t}(t,i)$$

$$\geq \lambda c(i, \widetilde{f}(t,i))e^{\lambda V^*(t,i)} + \sum_{j=0}^{j_n-1} \left( e^{\lambda V^*(t,j)} - e^{\lambda V^*(t,j_n)} \right) q(j|i, \widetilde{f}(t,i))$$

$$- e^{|\lambda|MT} d_H(A(i), A_n(i)) \left( |\lambda|L_i + 2\sum_{j=0}^{j_n-1} L_{ij} \right) - e^{|\lambda|MT}(\rho_1 + d_1 + R)\frac{w^2(i)}{w(j_n)}$$

$$\geq \lambda c(i, \widetilde{f}(t,i))e^{\lambda V^*(t,i)} + \sum_{j\in S_n} e^{\lambda V^*(t,j)} q_n(j|i, \widetilde{f}(t,i))$$

$$- \left[ \widetilde{M} + e^{|\lambda|MT}(\rho_1 + d_1 + R) \right] \frac{w^2(i)}{w(j_n)}. \tag{5.5}$$

Observe that (5.1), Assumptions 3.1 and 5.1 yield

$$\sum_{j\in S_n} w^l(j)q_n(j|i,a) \leq \rho_l w^l(i) + d_l \quad \text{and} \quad q_n^*(i) := \sup_{a\in A_n(i)} |q_n(i|i,a)| \leq Rw(i)$$

$$\tag{5.6}$$

for all $a \in A_n(i)$ and $l = 1, 2$. By (5.5), Theorems 3.1 and 4.1(b), we get

$$
E_n^{j,\widetilde{f}} \left[ \exp \left( \lambda \int_t^T c(\xi_s, \widetilde{f}(s, \xi_s)) ds \right) \Big| \xi_t = i \right] - e^{\lambda V^*(t,i)}
$$
$$
\leq \frac{\widetilde{M} + e^{|\lambda|MT}(\rho_1 + d_1 + R)}{w(j_n)}
$$
$$
\times E_n^{j,\widetilde{f}} \left[ \int_t^T \exp \left( \lambda \int_t^r c(\xi_v, \widetilde{f}(v, \xi_v)) dv \right) w^2(\xi_r) dr \Big| \xi_t = i \right]
$$

for all $j \in S_n$. Then it follows from (5.6), Lemma 6.3 in Guo and Hernández-Lerma (2009), Assumption 3.1(iii) and the last inequality that

$$
e^{\lambda V_n^{\widetilde{f}}(t,i)} - e^{\lambda V^*(t,i)} \leq \frac{\left[ \widetilde{M} + e^{|\lambda|MT}(\rho_1 + d_1 + R) \right] e^{|\lambda|MT}}{w(j_n)} \int_t^T E_n^{j,\widetilde{f}} \left[ w^2(\xi_r) \Big| \xi_t = i \right] dr
$$
$$
\leq \frac{\left[ \widetilde{M} + e^{|\lambda|MT}(\rho_1 + d_1 + R) \right] e^{|\lambda|MT}}{w(j_n)} \int_t^T \left[ e^{\rho_2(r-t)} w^2(i) \right.
$$
$$
\left. + \frac{d_2}{\rho_2} e^{\rho_2(r-t)} \right] dr
$$
$$
\leq \frac{\left[ \widetilde{M} + e^{|\lambda|MT}(\rho_1 + d_1 + R) \right] T e^{|\lambda|MT}}{w(j_n)} \left[ e^{\rho_2 T} + \frac{d_2}{\rho_2} e^{\rho_2 T} \right] w^2(i).
$$

Hence, letting $\widetilde{R} := T e^{|\lambda|MT} \left[ \widetilde{M} + e^{|\lambda|MT}(\rho_1 + d_1 + R) \right] \times \left[ e^{\rho_2 T} + \frac{d_2}{\rho_2} e^{\rho_2 T} \right]$, we have

$$
e^{\lambda V_n^*(t,i)} - e^{\lambda V^*(t,i)} \leq \frac{\widetilde{R} w^2(i)}{w(j_n)} \quad \text{for} \quad \lambda > 0. \tag{5.7}
$$

Following the similar arguments of (5.7), we can obtain

$$
\left| e^{\lambda V_n^*(t,i)} - e^{\lambda V^*(t,i)} \right| \leq \frac{\widetilde{R} w^2(i)}{w(j_n)} \quad \text{for} \quad \lambda \neq 0. \tag{5.8}
$$

Moreover, employing the differential mean value theorem and Theorem 4.1(a), we get

$$
\left| e^{\lambda V_n^*(t,i)} - e^{\lambda V^*(t,i)} \right| \geq |\lambda| e^{-|\lambda|MT} |V_n^*(t,i) - V^*(t,i)|,
$$

which together with (5.8) implies the assertion.                                                                       □

For each integer $m \geq 1$, we divide the interval $[0, T]$ into $m$ equal parts with the following discrete points: $t_0 := T$ and $t_l := t_0 - \frac{T}{m} l$ for all $l = 1, \ldots, m$. For each integer $n \geq 1$, define the iteration to compute approximately the optimal value as

follows:

$$W_m(t_l, i) := W_m(t_{l-1}, i) + sgn(\lambda)\frac{T}{m} \min_{a \in A_n(i)} sgn(\lambda)\Big\{\lambda c(i, a) W_m(t_{l-1}, i)$$

$$+ \sum_{j \in S_n} W_m(t_{l-1}, j) q_n(j|i, a)\Big\} \tag{5.9}$$

with $W_m(t_0, i) = 1$ for all $i \in S_n$ and $l = 1, \ldots, m$. For each $n \geq 1, m \geq 1, i \in S_n$ and $l \in \{1, \ldots, m\}$, denote by $\mathcal{D}_{n,l}(i)$ the set of all the minimizers attaining the minimum of (5.9), and by $\mathcal{O}_{n,m}$ the set of all the policies with the following form:

$$f_{n,m}(t, i) := \begin{cases} h_{n,m}(i), & \text{if } t \in [0, t_{m-1}], \\ h_{n,l}(i), & \text{if } t \in (t_l, t_{l-1}](l = 1, \ldots, m-1), \\ a^*, & \text{if } t > T, \end{cases} \tag{5.10}$$

where $h_{n,l}(i)$ belongs to $\mathcal{D}_{n,l}(i)$ and $a^* \in A_n(i)$ is arbitrarily fixed.

Below we give the error estimations on the approximate computations of the optimal value and an optimal policy via employing the iteration defined by (5.9).

**Theorem 5.2** *Let $R_1$ be as in Theorem 5.1. Under the conditions in Theorem 5.1, the following statements hold.*

(a) $\left|e^{\lambda V_n^*(t_l, i)} - W_m(t_l, i)\right| \leq \frac{TR_2}{m} w(j_n)\left(e^{T(|\lambda|M+2R)w(j_n)} - 1\right)$ *for all $n \geq 1$, $i \in S_n$, $m \geq 1$ and $l = 0, 1, \ldots, m$, where $R_2 = (|\lambda|M + 2R)e^{|\lambda|MT}$.*

(b) *For any $n \geq 1$ and $\tau \in (0, 1)$, there exists a positive integer $m_\tau$ such that*

$$(1 - \tau)e^{-|\lambda|MT} \leq W_m(0, i) \leq (1 + \tau)e^{|\lambda|MT} \quad and$$

$$\left|V^*(0, i) - \frac{1}{\lambda} \ln W_m(0, i)\right| \leq \frac{R_1 w^2(i)}{w(j_n)}$$

$$+ \frac{TR_2 e^{|\lambda|MT}}{m|\lambda|(1 - \tau)} w(j_n)\left(e^{T(|\lambda|M+2R)w(j_n)} - 1\right)$$

*for all $i \in S_n$ and $m \geq m_\tau$.*

(c) *For any $n \geq 1$, $m \geq m_\tau$ and $f_{n,m} \in \mathcal{O}_{n,m}$, we have*

$$\left|V^*(0, i) - V_n^{f_{n,m}}(0, i)\right| \leq \frac{R_1 w^2(i)}{w(j_n)}$$

$$+ \frac{2TR_2 e^{|\lambda|MT}}{m|\lambda|(1 - \tau)} w(j_n)\left(e^{T(|\lambda|M+2R)w(j_n)} - 1\right)$$

*for all $i \in S_n$.*

*Proof* (a) Fix any $n \geq 1$, $i \in S_n$, $m \geq 1$ and $l \in \{1, \ldots, m\}$. It follows from (5.1) and (5.6) that the conditions in Theorem 4.1 are satisfied for the control model $\mathcal{M}_n$. Then

by Theorem 4.1 we have $V_n^*(T, i) = 0$ and

$$- \lambda e^{\lambda V_n^*(t,i)} \frac{\partial V_n^*}{\partial t}(t, i)$$

$$= sgn(\lambda) \min_{a \in A_n(i)} sgn(\lambda) \left\{ \lambda c(i, a) e^{\lambda V_n^*(t,i)} + \sum_{j \in S_n} e^{\lambda V_n^*(t,j)} q_n(j|i, a) \right\} \quad (5.11)$$

for all $t \in [0, T]$. Thus, employing (5.6) and (5.11) we obtain

$$\left| e^{\lambda V_n^*(t,i)} - e^{\lambda V_n^*(s,i)} \right|$$

$$= \left| \int_t^s sgn(\lambda) \min_{a \in A_n(i)} sgn(\lambda) \left\{ \lambda c(i, a) e^{\lambda V_n^*(r,i)} + \sum_{j \in S_n} e^{\lambda V_n^*(r,j)} q_n(j|i, a) \right\} dr \right|$$

$$\leq (|\lambda| M + 2R) e^{|\lambda| M T} w(i) |t - s| \quad (5.12)$$

for all $s, t \in [0, T]$. Moreover, we have

$$\left| e^{\lambda V_n^*(t_l,i)} - W_m(t_l, i) \right|$$

$$= \left| e^{\lambda V_n^*(t_{l-1},i)} + \int_{t_l}^{t_{l-1}} sgn(\lambda) \min_{a \in A_n(i)} sgn(\lambda) \left\{ \lambda c(i, a) e^{\lambda V_n^*(t,i)} \right. \right.$$

$$+ \sum_{j \in S_n} e^{\lambda V_n^*(t,j)} q_n(j|i, a) \right\} dt$$

$$- W_m(t_{l-1}, i) - sgn(\lambda) \frac{T}{m} \min_{a \in A_n(i)} sgn(\lambda) \left\{ \lambda c(i, a) W_m(t_{l-1}, i) \right.$$

$$+ \left. \left. \sum_{j \in S_n} W_m(t_{l-1}, j) q_n(j|i, a) \right\} \right|$$

$$\leq \left| e^{\lambda V_n^*(t_{l-1},i)} - W_m(t_{l-1}, i) \right| + \int_{t_l}^{t_{l-1}} \max_{a \in A_n(i)} \left| \lambda c(i, a) \left( e^{\lambda V_n^*(t,i)} - W_m(t_{l-1}, i) \right) \right| dt$$

$$+ \int_{t_l}^{t_{l-1}} \max_{a \in A_n(i)} \left| \sum_{j \in S_n} \left( e^{\lambda V_n^*(t,j)} - W_m(t_{l-1}, j) \right) q_n(j|i, a) \right| dt$$

$$\leq \left| e^{\lambda V_n^*(t_{l-1},i)} - W_m(t_{l-1}, i) \right| + |\lambda| M \int_{t_l}^{t_{l-1}} \left| e^{\lambda V_n^*(t,i)} - e^{\lambda V_n^*(t_{l-1},i)} \right| dt$$

$$+ \frac{T}{m} |\lambda| M \left| e^{\lambda V_n^*(t_{l-1},i)} - W_m(t_{l-1}, i) \right| + \int_{t_l}^{t_{l-1}} \max_{a \in A_n(i)} \left| \sum_{j \in S_n} \left( e^{\lambda V_n^*(t,j)} \right. \right.$$

$$-e^{\lambda V_n^*(t_{l-1},j)}\Big) q_n(j|i,a)\Big| dt$$

$$+\frac{T}{m}\max_{a\in A_n(i)}\left|\sum_{j\in S_n}\left(e^{\lambda V_n^*(t_{l-1},j)}-W_m(t_{l-1},j)\right)q_n(j|i,a)\right|$$

$$\leq \left|e^{\lambda V_n^*(t_{l-1},i)}-W_m(t_{l-1},i)\right|+\left(\frac{T}{m}\right)^2(|\lambda|M+2R)^2 e^{|\lambda|MT}w^2(j_n)$$

$$+\frac{T}{m}(|\lambda|M+2R)w(i)\max_{j\in S_n}\left|e^{\lambda V_n^*(t_{l-1},j)}-W_m(t_{l-1},j)\right|,$$

where the last inequality follows from (5.6) and (5.12). Hence, we obtain

$$\max_{j\in S_n}\left|e^{\lambda V_n^*(t_l,j)}-W_m(t_l,j)\right|$$

$$\leq \left[1+\frac{T}{m}(|\lambda|M+2R)w(j_n)\right]\max_{j\in S_n}\left|e^{\lambda V_n^*(t_{l-1},j)}-W_m(t_{l-1},j)\right|$$

$$+\left(\frac{T}{m}\right)^2(|\lambda|M+2R)^2 e^{|\lambda|MT}w^2(j_n)$$

$$=\left[1+\frac{T}{m}(|\lambda|M+2R)w(j_n)\right]\times\left[\max_{j\in S_n}\left|e^{\lambda V_n^*(t_{l-1},j)}-W_m(t_{l-1},j)\right|\right.$$

$$\left.+\frac{T}{m}(|\lambda|M+2R)e^{|\lambda|MT}w(j_n)\right]-\frac{T}{m}(|\lambda|M+2R)e^{|\lambda|MT}w(j_n).$$

Employing the last inequality and the induction, we get

$$\max_{j\in S_n}\left|e^{\lambda V_n^*(t_l,j)}-W_m(t_l,j)\right|\leq\frac{T}{m}(|\lambda|M+2R)e^{|\lambda|MT}w(j_n)\left(e^{\frac{T}{m}l(|\lambda|M+2R)w(j_n)}-1\right)$$

$$\leq\frac{T}{m}(|\lambda|M+2R)e^{|\lambda|MT}w(j_n)\times\left(e^{T(|\lambda|M+2R)w(j_n)}-1\right).$$

Therefore, the assertion holds.

(b) Fix any $n\geq 1$ and $i\in S_n$. By part (a), we have $\lim_{m\to\infty}W_m(0,i)=e^{\lambda V_n^*(0,i)}$, which implies that for any $\tau\in(0,1)$, there exists a positive integer $m_\tau$ satisfying

$$e^{\lambda V_n^*(0,i)}-\tau e^{-|\lambda|MT}\leq W_m(0,i)\leq e^{\lambda V_n^*(0,i)}+\tau e^{-|\lambda|MT}$$

for all $m\geq m_\tau$. Thus, employing the last inequalities we obtain

$$(1-\tau)e^{-|\lambda|MT}\leq W_m(0,i)\leq(1+\tau)e^{|\lambda|MT}\quad\text{for all}\quad m\geq m_\tau. \qquad (5.13)$$

Below we discuss the cases $\lambda>0$ and $\lambda<0$ separately. Note that

$$\frac{1}{\lambda}\ln(1+\tau)\leq\frac{1}{\lambda}\ln\frac{1}{1-\tau}\text{ for }\lambda>0\quad\text{and}\quad\frac{1}{\lambda}\ln(1+\tau)\geq\frac{1}{\lambda}\ln\frac{1}{1-\tau}\text{ for }\lambda<0.$$

$$(5.14)$$

Case 1: $\lambda > 0$. For each $m \geq m_\tau$, by (5.13) and (5.14) we get

$$-MT + \frac{1}{\lambda}\ln(1-\tau) \leq \frac{1}{\lambda}\ln W_m(0,i) \leq MT + \frac{1}{\lambda}\ln(1+\tau) \leq MT - \frac{1}{\lambda}\ln(1-\tau).$$

Case 2: $\lambda < 0$. For each $m \geq m_\tau$, employing (5.13) and (5.14) we have

$$-MT - \frac{1}{\lambda}\ln(1-\tau) \leq -MT + \frac{1}{\lambda}\ln(1+\tau) \leq \frac{1}{\lambda}\ln W_m(0,i) \leq MT + \frac{1}{\lambda}\ln(1-\tau).$$

Hence, we obtain

$$-MT + \frac{1}{|\lambda|}\ln(1-\tau) \leq \frac{1}{\lambda}\ln W_m(0,i) \leq MT - \frac{1}{|\lambda|}\ln(1-\tau) \qquad (5.15)$$

for all $m \geq m_\tau$ and $\lambda \neq 0$. Moreover, it follows from the differential mean value theorem and (5.15) that

$$\left| e^{\lambda V_n^*(0,i)} - W_m(0,i) \right| \geq |\lambda|(1-\tau)e^{-|\lambda|MT}\left| V_n^*(0,i) - \frac{1}{\lambda}\ln W_m(0,i) \right|,$$

which together with part (a) implies

$$\left| V_n^*(0,i) - \frac{1}{\lambda}\ln W_m(0,i) \right| \leq \frac{T}{m|\lambda|(1-\tau)}$$
$$\times (|\lambda|M + 2R)\, e^{2|\lambda|MT} w(j_n) \left( e^{T(|\lambda|M+2R)w(j_n)} - 1 \right)$$

for all $m \geq m_\tau$. Therefore, the desired result follows from the last inequality, Theorem 5.1 and the inequality $\left| V^*(0,i) - \frac{1}{\lambda}\ln W_m(0,i) \right| \leq \left| V^*(0,i) - V_n^*(0,i) \right| + \left| V_n^*(0,i) - \frac{1}{\lambda}\ln W_m(0,i) \right|$.

(c) Fix any $n \geq 1, m \geq m_\tau$ and $f_{n,m} \in \mathcal{O}_{n,m}$. Define an operator $\widetilde{\Gamma}$ on $B([0,T] \times S_n)$ as follows:

$$\widetilde{\Gamma}g(t,i) := e^{\bar{\alpha}t} + e^{\bar{\alpha}t} \int_t^T e^{-\bar{\alpha}s}\left( \lambda c(i, f_{n,m}(s,i))g(s,i) \right.$$
$$\left. + \sum_{j \in S_n} g(s,j)q_n(j|i, f_{n,m}(s,i)) \right) ds$$

for all $g \in B([0,T] \times S_n)$, where $\bar{\alpha}$ is an arbitrary positive constant. Then by (5.6) and Assumption 5.1(i) we have $|\widetilde{\Gamma}g(t,i)| \leq e^{\bar{\alpha}T} + Te^{\bar{\alpha}T}(|\lambda|M + 2Rw(j_n))\|g\|$ for all $(t,i) \in [0,T] \times S_n$, which yields that $\widetilde{\Gamma}$ is a map from $B([0,T] \times S_n)$ into itself. Moreover, for any $g_1, g_2 \in B([0,T] \times S_n)$, direct calculations give

$$|\widetilde{\Gamma}g_1(t,i) - \widetilde{\Gamma}g_2(t,i)| \leq \frac{|\lambda|M + 2Rw(j_n)}{\bar{\alpha}}\|g_1 - g_2\|$$

for all $(t, i) \in [0, T] \times S_n$. Hence, letting $\overline{\alpha} = |\lambda| M + 2Rw(j_n) + 1$, we obtain that $\widetilde{\Gamma}$ is a contraction operator on $B([0, T] \times S_n)$. Therefore, the Banach fixed point theorem implies the existence of a function $\widetilde{g} \in B([0, T] \times S_n)$ satisfying

$$
\begin{aligned}
\widetilde{g}(t, i) = e^{\overline{\alpha}t} + e^{\overline{\alpha}t} \int_t^T e^{-\overline{\alpha}s} \bigg( & \lambda c(i, f_{n,m}(s, i)) \widetilde{g}(s, i) \\
& + \sum_{j \in S_n} \widetilde{g}(s, j) q_n(j|i, f_{n,m}(s, i)) \bigg) ds
\end{aligned}
\tag{5.16}
$$

for all $(t, i) \in [0, T] \times S_n$ and $\overline{\alpha} = |\lambda| M + 2Rw(j_n) + 1$. Set $\widetilde{u}(t, i) := e^{-(|\lambda| M + 2Rw(j_n) + 1)t} \widetilde{g}(t, i)$. Then by (5.16) we have

$$
\widetilde{u}(t, i) = 1 + \int_t^T \bigg( \lambda c(i, f_{n,m}(s, i)) \widetilde{u}(s, i) + \sum_{j \in S_n} \widetilde{u}(s, j) q_n(j|i, f_{n,m}(s, i)) \bigg) ds
\tag{5.17}
$$

for all $(t, i) \in [0, T] \times S_n$. Thus, the last equality yields that for each $i \in S_n$, $\widetilde{u}(\cdot, i)$ is absolutely continuous on $[0, T]$ and

$$
-\frac{\partial \widetilde{u}}{\partial t}(t, i) = \lambda c(i, f_{n,m}(t, i)) \widetilde{u}(t, i) + \sum_{j \in S_n} \widetilde{u}(t, j) q_n(j|i, f_{n,m}(t, i))
\tag{5.18}
$$

for a.e. $t \in [0, T]$. Moreover, from the proof of Theorem 3.1, we conclude that the Feynman–Kac formula is also applicable to the function $\widetilde{u}$. Hence, employing (5.18) and following the similar arguments of (4.6), we have $\widetilde{u}(t, i) = e^{\lambda V_n^{f_{n,m}}(t,i)}$ for all $(t, i) \in [0, T] \times S_n$. Therefore, by (5.17) we get

$$
\begin{aligned}
e^{\lambda V_n^{f_{n,m}}(t,i)} = 1 + \int_t^T \bigg( & \lambda c(i, f_{n,m}(s, i)) e^{\lambda V_n^{f_{n,m}}(s,i)} \\
& + \sum_{j \in S_n} e^{\lambda V_n^{f_{n,m}}(s,j)} q_n(j|i, f_{n,m}(s, i)) \bigg) ds
\end{aligned}
$$

for all $(t, i) \in [0, T] \times S_n$. Then using the last equality and following the same techniques in the proofs of parts (a) and (b), we obtain

$$
\left| V_n^{f_{n,m}}(0, i) - \frac{1}{\lambda} \ln W_m(0, i) \right| \leq \frac{T R_2 e^{|\lambda| MT}}{m |\lambda| (1 - \tau)} w(j_n) \left( e^{T(|\lambda| M + 2R) w(j_n)} - 1 \right)
$$

for all $i \in S_n$, which together with part (b) gives the assertion. $\qquad \square$

*Remark 5.1* (a) Theorems 5.1 and 5.2 are new for the risk-sensitive finite-horizon cost criterion. The iteration defined by (5.9) provides a numerical method to compute approximately the optimal value and an optimal policy of the model $\mathcal{M}$.

(b) For the control model $\mathcal{M}$ with finite states and finite actions, there exists a positive constant $\widehat{R}$ such that $q^*(i) \leq \widehat{R}$ for all $i \in S$. Moreover, we do not need to construct a sequence of the control models $\{\mathcal{M}_n, n \geq 1\}$. Thus, employing the same technique used in the proof of Theorem 5.2, we obtain that for any $\tau \in (0, 1)$, there exists a positive integer $m_\tau$ such that

$$\left| V^*(0, i) - \frac{1}{\lambda} \ln W_m(0, i) \right| \leq \frac{T}{m|\lambda|(1 - \tau)} (|\lambda|M$$
$$+ 2\widehat{R}) e^{2|\lambda|MT} \left( e^{T(|\lambda|M + 2\widehat{R})} - 1 \right) \text{ and}$$

$$\left| V^*(0, i) - V^{f_m}(0, i) \right| \leq \frac{2T}{m|\lambda|(1 - \tau)} (|\lambda|M$$
$$+ 2\widehat{R}) e^{2|\lambda|MT} \left( e^{T(|\lambda|M + 2\widehat{R})} - 1 \right)$$

for all $i \in S$ and $m \geq m_\tau$, where the policy $f_m$ is as in (5.10). Hence, the last two inequalities yield that the accuracy of the approximation given by (5.9) and (5.10) is of order $m^{-1}$.

## 6 An example

In this section, we illustrate the application of our main results with a controlled birth and death system and use the iteration in Sect. 5 to compute approximately the optimal value.

*Example 6.1* [*A controlled birth and death system in* Guo and Hernández-Lerma (2009)] Consider a controlled birth and death system in which the state variable represents the population size. Let the positive constants $\beta$ and $\gamma$ denote the natural birth and death rates, respectively. Suppose that the immigration parameter denoted by $a_1$ and the emigration parameter denoted by $a_2$ can be controlled by a decision-maker. When the population size of the system is not less than one, the decision-maker chooses an immigration parameter from a given set $[0, \kappa]$ ($\kappa > 0$) and an emigration parameter from a given set $[\zeta_1, \zeta_2]$ ($\zeta_2 > \zeta_1 > 0$) to control the population size. When the population size equals zero, the decision-maker only needs to choose an immigration parameter from the set $[0, \kappa]$ and it is natural to take $a_2 \equiv 0$. Moreover, we assume that there exists a positive integer $i^*$ such that the cost of regulating the system is too high when the population size exceeds the integer $i^*$. Thus, we suppose that the cost takes a large enough positive value $Q$ when the population size is greater than $i^*$. If the population size is $i \in \{0, 1, 2, \ldots, i^*\}$, we suppose that this state and the action $(a_1, a_2)$ incur a cost $(|a_1 - \eta_1| + |a_2 - \eta_2|)i$, where $\eta_1$ and $\eta_2$ are given positive constants.

Now we formulate the above controlled birth and death system as a CTMDP with the state space given by $S = \{0, 1, 2 \ldots\}$, the sets of all admissible actions given by $A(0) = [0, \kappa] \times \{0\}$ and $A(i) = [0, \kappa] \times [\zeta_1, \zeta_2]$ for all $i \geq 1$, the transition rate given by $q(1|0, (a_1, 0)) = -q(0|0, (a_1, 0)) = a_1$ for all $a_1 \in [0, \kappa]$ and

$$q(j|i, a) = \begin{cases} \beta i + a_1, & \text{if } j = i + 1, \\ -(\beta + \gamma)i - a_1 - a_2, & \text{if } j = i, \\ \gamma i + a_2, & \text{if } j = i - 1, \\ 0, & \text{otherwise,} \end{cases}$$

for all $i \geq 1$ and $a = (a_1, a_2) \in A(i)$, and the cost given by

$$c(i, a) = \begin{cases} (|a_1 - \eta_1| + |a_2 - \eta_2|)i, & \text{if } 0 \leq i \leq i^*, \\ Q, & \text{if } i > i^*, \end{cases}$$

for all $a = (a_1, a_2) \in A(i)$.

Then we have the following statement.

**Proposition 6.1** *The controlled birth and death system in Example* 6.1 *satisfies Assumptions* 3.1, 3.2 *and* 5.1. *Hence, (by Theorem* 4.1*), there exists an optimal deterministic Markov policy.*

*Proof* Take $w(i) = i + 1$ for all $i \in S$. Then Assumption 5.1(i) holds. Moreover, by the description of the model, we obtain

$$q^*(i) \leq (\beta + \gamma)i + \kappa + \zeta_2 \leq \max\{\beta + \gamma, \kappa + \zeta_2\}w(i),$$

$$\sum_{j \in S} w(j)q(j|i, a) = (\beta - \gamma)i + a_1 - a_2 \leq \beta w(i) + \kappa,$$

$$\sum_{j \in S} w^2(j)q(j|i, a) = 2(\beta - \gamma)i^2 + (3\beta - \gamma + 2a_1 - 2a_2)i + 3a_1 - a_2$$

$$\leq (5\beta + 2\kappa)w^2(i) + 3\kappa$$

for all $i \geq 1$ and $a = (a_1, a_2) \in A(i)$, and

$$q^*(0) \leq \kappa w(0) \leq \max\{\beta + \gamma, \kappa + \zeta_2\}w(0),$$

$$\sum_{j \in S} w(j)q(j|0, a) = a_1 \leq \beta w(0) + \kappa,$$

$$\sum_{j \in S} w^2(j)q(j|0, a) = 3a_1 \leq (5\beta + 2\kappa)w^2(0) + 3\kappa$$

for all $a = (a_1, a_2) \in A(0)$. Hence, Assumptions 3.1(i), (ii), 5.1(ii) are satisfied with $\rho_1 = \beta$, $d_1 = \kappa$, $\rho_2 = 5\beta + 2\kappa$, $d_2 = 3\kappa$ and $R = \max\{\beta + \gamma, \kappa + \zeta_2\}$. Moreover, we have $|c(i, a)| \leq \max\{(\kappa + \zeta_2 + \eta_1 + \eta_2)i^*, Q\}$. Thus, Assumption 3.1(iii) holds with $M = \max\{(\kappa + \zeta_2 + \eta_1 + \eta_2)i^*, Q\}$. Furthermore, direct calculations give $|c(i, a) - c(i, b)| \leq (|a_1 - b_1| + |a_2 - b_2|)i$ for all $0 \leq i \leq i^*$, $|c(i, a) - c(i, b)| \leq |a_1 - b_1| + |a_2 - b_2|$ for all $i > i^*$ and $|q(j|i, a) - q(j|i, b)| \leq |a_1 - b_1| + |a_2 - b_2|$ for all $i, j \in S$ and $a = (a_1, a_2), b = (b_1, b_2) \in A(i)$, which together with the description of the model imply that Assumptions 3.2 and 5.1(iii) are satisfied with $L_i = i$ for $0 \leq i \leq i^*$, $L_i = 1$ for $i > i^*$ and $L_{ij} = 1$. Therefore, we complete the proof of the proposition. $\square$
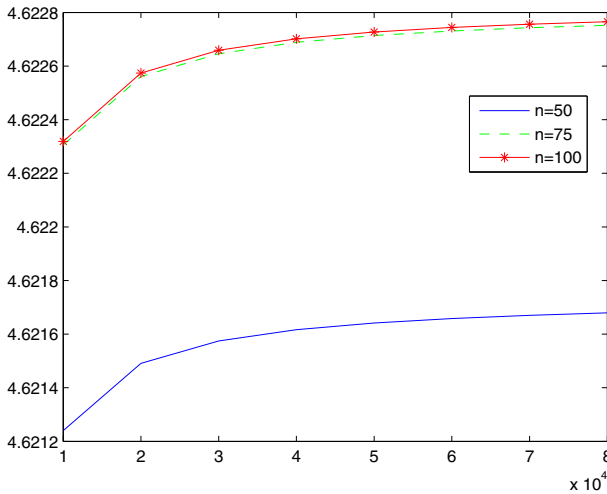
**Fig. 1** Value $\ln W_m^{10}(0, 3)$ of the model $\mathcal{M}_n$ for $n = 50, 75, 100$ and $m = 10,000, \ldots, 80,000$

For each $n \geq 1$, the control model $\mathcal{M}_n$ is given by $S_n = \{0, 1, \ldots, n\}$, $A_n(0) = \{\frac{\kappa l}{n^2} : l = 0, 1, \ldots, n^2\} \times \{0\}$, $A_n(i) = \{\frac{\kappa l}{n^2} : l = 0, 1, \ldots, n^2\} \times \{\zeta_1 + \frac{(\zeta_2 - \zeta_1)l}{n^2} : l = 0, 1, \ldots, n^2\}$ for all $i = 1, 2, \ldots, n$. Then for each $n \geq 1$ and $i \in S_n$, direct calculations give

$$d_H(A(i), A_n(i)) \leq \frac{\kappa + \zeta_2 - \zeta_1}{n^2} \leq \frac{2e^{|\lambda|MT}(|\lambda|i^* + 2)(\kappa + \zeta_2 - \zeta_1)w^2(i)}{e^{|\lambda|MT}\left(|\lambda|L_i + 2\sum_{j=0}^{n-1} L_{ij}\right)w(n)}.$$

Hence, (5.2) in Theorem 5.1 holds with $\widetilde{M} = 2e^{|\lambda|MT}(|\lambda|i^* + 2)(\kappa + \zeta_2 - \zeta_1)$.

For a numerical experimentation of Example 6.1, we take the following values of the parameters: $T = 5$, $\lambda = 0.1$, $\beta = 0.9$, $\gamma = 1$, $\kappa = 1$, $\zeta_1 = 0.4$, $\zeta_2 = 1$, $\eta_1 = 1.1$, $\eta_2 = 1.2$, $i^* = 10,000$, $Q = 1,000,000$. For $n = 50, 75, 100$ and $m = 10,000, \ldots, 80,000$, via the iteration defined by (5.9), we obtain the values of $\ln W_m^{10}(0, 3)$ as displayed in Fig. 1. Empirically, the convergence is faster than that given in Theorem 5.2. This is due to the fact that the bounds used to obtain the error estimations in Theorem 5.2 are very conservative. Moreover, as can be seen in Fig. 1, the value of $V^*(0, 3)$ approximately equals 4.62277.

## 7 Concluding remarks

In this paper we have studied the risk-sensitive finite-horizon cost criterion for CTMDPs with the denumerable states, bounded cost rates and possibly unbounded transition rates. The risk-sensitivity coefficient can take any nonzero value. Under the suitable conditions, we have established the existence and uniqueness of the solution to the optimality equation and shown the existence of an optimal deterministic Markov policy. Moreover, we have proposed a tractable numerical method for the

approximate computations of an optimal policy and the optimal value, and obtained the corresponding error estimations. It should be mentioned that the Feynman–Kac formula plays a crucial role in the study of the risk-sensitive finite-horizon cost criterion. The Feynman–Kac formula in Theorem 3.1 is only applicable to the case of the bounded cost rates and bounded optimal value function. For the case of the unbounded cost rates, the corresponding optimal value function is also unbounded. Hence, in order to study the risk-sensitive finite-horizon cost criterion with the unbounded cost rates, we need to extend the Feynman–Kac formula in Theorem 3.1 to the case which is applicable to the unbounded cost rates and unbounded optimal value function. How to extend the Feynman–Kac formula remains an open problem.

# References

Cavazos-Cadena R, Hernández-Hernández D (2011) Discounted approximations for risk-sensitive average criteria in Markov decision chains with finite state space. Math Oper Res 36:133–146

Confortola F, Fuhrman M (2014) Backward stochastic differential equations associated to jump Markov processes and applications. Stoch Process Appl 124:289–316

Di Masi GB, Stettner L (2007) Infinite horizon risk sensitive control of discrete time Markov processes under minorization property. SIAM J Control Optim 46:231–252

Ghosh MK, Saha S (2014) Risk-sensitive control of continuous time Markov chains. Stochastics 86:655–675

Guo XP, Hernández-Lerma O (2009) Continuous-time Markov decision processes: theory and applications. Springer, Berlin

Guo XP, Zhang WZ (2014) Convergence of controlled models and finite-state approximation for discounted continuous-time Markov decision processes with constraints. Eur J Oper Res 238:486–496

Guo XP, Huang XX, Huang YH (2015) Finite horizon optimality for continuous-time Markov decision processes with unbounded transition rates. Adv Appl Probab 47:1064–1087

Hernández-Lerma O, Lasserre JB (1999) Further topics on discrete-time Markov control processes. Springer, New York

Jaśkiewicz A (2007) Average optimality for risk-sensitive control with general state space. Ann Appl Probab 17:654–675

Kitaev MY, Rykov VV (1995) Controlled queueing systems. CRC Press, Boca Raton

Miller BL (1968) Finite state continuous time Markov decision processes with finite planning horizon. SIAM J Control 6:266–280

Prieto-Rumeau T, Hernández-Lerma O (2012) Discounted continuous-time controlled Markov chains: convergence of control models. J Appl Probab 49:1072–1090

Prieto-Rumeau T, Lorenzo JM (2010) Approximating ergodic average reward continuous-time controlled Markov chains. IEEE Trans Autom Control 55:201–207

Puterman ML (1994) Markov decision processes: discrete stochastic dynamic programming. Wiley, New York

Royden HL (1988) Real analysis. Macmillan, New York

van Dijk NM (1988) On the finite horizon Bellman equation for controlled Markov jump models with unbounded characteristics: existence and approximation. Stoch Process Appl 28:141–157

van Dijk NM (1989) A note on constructing $\varepsilon$-optimal policies for controlled Markov jump models with unbounded characteristics. Stochastics 27:51–58

Wei QD, Chen X (2014) Strong average optimality criterion for continuous-time Markov decision processes. Kybernetika 50:950–977