Discrete Event Dynamic Systems (2019) 29:445–471
https://doi.org/10.1007/s10626-019-00292-y

Risk-sensitive continuous-time Markov decision processes with unbounded rates and Borel spaces

Xianping Guo[1,2] · Junyu Zhang[1,2]

Received: 15 November 2018 / Accepted: 5 September 2019 / Published online: 19 October 2019

Abstract

This paper considers the finite-horizon *risk-sensitive* optimality for continuous-time Markov decision processes, and focuses on the more general case that the transition rates are unbounded, cost/reward rates are allowed to be unbounded from below and from above, the policies can be history-dependent, and the state and action spaces are Borel ones. Under mild conditions imposed on the decision process's *primitive data*, we establish the existence of a solution to the corresponding optimality equation (OE) by a so called approximation technique. Then, using the OE and the *extension of Dynkin's formula* developed here, we prove the existence of an optimal Markov policy, and verify that the value function is the unique solution to the OE. Finally, we give an example to illustrate the difference between our conditions and those in the previous literature.

Keywords Continuous-time Markov decision process · Finite horizon risk-sensitive criterion · History-dependent policy · Unbounded transition/cost rates · Optimal policy

1 Introduction

Continuous-time Markov decision processes (CTMDPs) are an important class of stochastic optimality control problems and have been widely studied; see, for instance, the monographs (Guo and Hernández-Lerma 2009; Prieto-Rumeau and Hernández-Lerma 2012) and the extensive references therein. In most existing literature on CTMDPs, the infinite-horizon expected discounted criterion (Guo 2007; Guo and Hernández-Lerma 2009; Guo and Song 2011; Guo and Piunovskiy 2011; Piunovskiy and Zhang 2011; Prieto-Rumeau and Hernández-Lerma 2012), the long-run expected average criterion (Guo and Hernández-Lerma 2009; Guo et al. 2012; Prieto-Rumeau and Hernández-Lerma 2012; Wei and Chen 2017; Xia 2014), and the finite-horizon expected criterion (Guo et al. 2015b; Yushkevich 1977), are the commonly used optimality criteria. All these expected criteria are risk-neutral and cannot reflect the attitude of a decision-maker to the risk. Since many decision-makers

✉ Junyu Zhang
mcszhjy@mail.sysu.edu.cn

[1] School of Mathematics, Sun Yat-Sen University, Guangzhou, 510275, China

[2] Guangdong Province Key Laboratory of Computational Science, Guangzhou, China

in the real-world applications may be either risk-seeking or risk-averse, in order to take the risk-sensitivity of a decision-maker into an optimality criterion, the risk-sensitive criteria have been employed and studied in CTMDPs. In this paper, we will further study a risk-sensitive criterion in CTMDPs and focus on the *finite horizon* case. Thus we shall pinpoint neither the main results in the earlier literature on the infinite-horizon risk-sensitive discounted criterion and the long-run risk-sensitive average criterion in Ghosh and Saha (2014), Huo et al. (2017), Kumar and Chandan (2013), and Zhang (2017) and Ghosh and Saha (2014), Kumar and Chandan (2015), and Kumar and Chandan (2013) respectively for CTMDPs, nor those on the risk-sensitive discrete-time Markov decision processes (Anantharam and Borkar 2017; Baüerle and Rieder 2014; Cavazos-Cadena and Hernndez-Hernndez 2011; Jaskiewicz 2007; Xia 2018) as well as on zero-sum risk-sensitive stochastic games (Basu and Ghosh 2014; Baüerle and Rieder 2017). Our concern is on the finite horizon risk-sensitive criterion for CTMDPs (Ghosh and Saha 2014; Wei 2016), which leads to that the value function also depends on both time and states, while the value functions for the infinite-horizon risk-sensitive discounted and average criteria in Ghosh and Saha (2014), Kumar and Chandan (2015), Kumar and Chandan (2013), and Zhang (2017) are independent of time. Therefore, the optimality equation for the finite horizon case (Ghosh and Saha 2014; Wei 2016) is rather different from that for the infinite horizon cases (Ghosh and Saha 2014; Kumar and Chandan 2013; Zhang 2017).

To the best of our knowledge, the finite horizon risk-sensitive criterion for CTMDPs is addressed only in Ghosh and Saha (2014), Huang (2018), and Wei (2016). Precisely, Ghosh and Saha (2014) uses the exponential utility function to characterize the risk-sensitivity of a decision-maker and obtains the existence of optimal Markov policies for the case of denumerable states, uniformly bounded transition rates, bounded cost rates, and the class of Markov policies. Wei (2016) extends the main results in Ghosh and Saha (2014) to the case of unbounded transition rates, establishes the existence of a solution to the corresponding optimality equation (OE) by the approximation technique in Guo et al. (2015b), and gives the corresponding error estimations of the approximations. However, the arguments in Ghosh and Saha (2014) and Wei (2016) need the assumptions of bounded cost rates and denumerable states, and the policies in Ghosh and Saha (2014) and Wei (2016) are independent of histories.

Huang (2018) also studies the finite horizon CTMDPs with the Borel state and action spaces and unbounded reward functions and transition rates. But the optimization criteria in Huang (2018) are very different from ours. The optimization criteria in Huang (2018) are firstly mean maximization, then for fixed mean, variance minimization. This is suitable for a risk-averse decision maker. The optimization criterion in this paper is risk-sensitivity, which is maximization of the exponential utility function. From Eq. 2.7 in Section 2, we can see that the exponential utility function is related to the mean and variance. Since the coefficient of variance is positive, our optimization criterion is suitable to a risk-seeking decision maker.

As indicated above, the finite horizon risk-sensitive criterion of for CTMDPs with unbounded cost/transition rates, Borel state and action spaces as well as randomized history-dependent policies, has *not* been studied yet, and it will be considered in this paper. Precisely, we study the CTMDPs having the following features: (1) the transition rates can be *unbounded*; (2) the cost rates are allowed to be *unbounded from below and from above*. (As the cost rates are allowed to take positive and negative values, they can be interpreted as reward rates rather than "cost rates" only); (3) the state and action spaces are Borel ones; (4) the policies can be randomized and *history-dependent*; and (5) the optimality criterion is *finite horizon* risk-sensitive.

First, we give a suitable condition, under which we establish the finiteness of the finite-horizon risk-sensitive criterion with unbounded cost rates using the Jensen inequality. This condition is new and satisfied for the bounded costs in Ghosh and Saha (2014) and Wei (2016); see Lemma 3.1.

Second, under the conditions slightly weaker than those in Guo and Hernández-Lerma (2009), Guo et al. (2012), Guo and Piunovskiy (2011), Piunovskiy and Zhang (2011), and Prieto-Rumeau and Hernández-Lerma (2012) on infinite horizon risk-neutral CTMDPs, we derive an extension of Dynkin's formula, which is a generalization of the *analog* of Ito-Dynkin's formula recently developed in Guo et al. (2015b) for the underlying state process $\{x_t, t \geq 0\}$ induced by the transition rates and *randomized history-dependent* policies (see Theorem 3.1 below). This result is also a natural extension of the Feynman-Kac formula in Wei (2016) for the continuous-time jump Markov process to the case of continuous-time jump "non-Markov" processes and a more larger class of functions $\varphi(\omega, t, x)$ of samples ω, time t and states x. On the one hand, since the analog of Ito-Dynkin's formula in Guo et al. (2015b) is designed for the forms of functions $\varphi(t, x)$ of time t and states x, it is not suitable for the more general forms of $\varphi(\omega, t, x)$ with an additional sample variable ω such as the function $e^{\int_0^t \int_A c(x_s(\omega),a)\pi(da|\omega,s)ds}\varphi(t, x)$ of (ω, t, x), which need to be considered for our case of history-dependent policies $\pi(da|\omega, s)$ and the cost rates $c(x, a)$ in the risk-sensitive CTMDPs; on the other hand, the Feynman-Kac formula in Wei (2016) is for Markov processes, and thus it is not applicable to the case that the underlying processes $\{x_t\}$ here may not be Markovian. Since our optimality problem is on the risk-sensitive criterion over the class of history-dependent policies, the two facts just mentioned above motivate our study on the *extension of the Dynkin's formula*.

Third, under suitable conditions as in Ghosh and Saha (2014), Guo and Hernández-Lerma (2009), Guo et al. (2012), Guo and Piunovskiy (2011), Piunovskiy and Zhang (2011), and Prieto-Rumeau and Hernández-Lerma (2012) on risk-neutral CTMDPs with infinite horizon, we prove the existence and uniqueness of a solution to the OE as well as the existence of an optimal Markov policy for the finite horizon risk-sensitive CTMDPs in three steps. The first step is to consider the simple case of bounded transition rates and bounded cost rates, and establishes the existence of a solution to the OE by the Banach's fixed point theorem, and also proves the existence of an optimal Markov policy; see Proposition 4.2. The second step is to deal with the case of unbounded transition rates and nonnegative cost rates. By constructing a sequence of the models of CTMDPs with bounded transition rates and bounded cost rates, using the results in the first step and some new properties of the value function of the finite horizon risk-sensitive CTMDPs, we prove that the limit of the sequence of the value functions of models of CTMDPs is a solution to the OE for the case of unbounded transition rates and nonnegative cost rates; see Proposition 4.3. In the third step, by designing a technique of approximations from nonnegative but unbounded cost rates to a more general case of cost rates being unbounded from above and from below, we further show the existence of a solution to the OE for the most general case of unbounded transition and cost rates; see Theorem 4.1. It should be mentioned that our approximation technique here is rather different from the approximation in Guo et al. (2015b) and Wei (2016) for denumerable states.

Fourth, using the existence of a solution to the OE and the *extension of the Dynkin's formula* developed here, we prove the existence of an optimal Markov policy, and also show that the value function is the unique solution to the optimality equation; see Theorem 4.1. All arguments here are direct and closed.

Finally, our conditions in this paper are an generalization of those in Ghosh and Saha (2014) and Wei (2016) on the finite horizon risk-sensitive CTMDPs. In order to further illustrate our main results and show the difference between the conditions in this paper and those in Ghosh and Saha (2014) and Wei (2016), we present an applied example, in which our conditions are satisfied, in which the transition and cost rates are both *unbounded*, in which the state and action spaces are non-denumerable, and for which some of the conditions in Ghosh and Saha (2014) and Wei (2016) fail to hold.

The rest of the paper is organized as follows. In Section 2 we introduce the optimality problem for the risk-sensitive CTMDPs. The main results are presented in Section 4 after giving technical preliminaries in Section 3, and are illustrated with an example in Section 5.

2 The optimal control problems

Notation For any Borel space X endowed with the Borel σ-algebra $\mathcal{B}(X)$, we will denote by $E^c := X \setminus E$ the complement of a subset E of X, by I_E the indicator function on any set E, by $\delta_z(dx)$ the Dirac measure at point $z \in X$ (i.e., $\delta_z(D) = I_D(z)$ for all $D \in \mathcal{B}(X)$), by $\mathbb{B}_1(X)$ the set of all bounded Borel measurable functions φ on X with the norm $\|\varphi\| := \sup_{x\in X} |\varphi(x)|$, and by $\mathcal{U}(X)$ the universal σ-algebra on X, that is, $\mathcal{U}(X) := \cap_{p\in P(X)}\mathcal{B}_p(X)$, where $P(X)$ represents the set of all probability measures on $\mathcal{B}(X)$, and $\mathcal{B}_p(X)$ is the completion of $\mathcal{B}(X)$ with respect to $p \in P(X)$. To discern the "measurability", we will say "Borel measurable" or "universally measurable" in the following.

The model of CTMDPs is a five-tuple

$$\mathbb{M} := \{S, A, (A(x), x \in S), c(x,a), q(dy|x,a)\}, \tag{2.1}$$

consisting of the following elements:

(a) a Borel space S, called the state space, whose elements are referred to as states of a system.

(b) a Borel space A, called the action space, whose elements are referred to as actions (or decisions) of a decision-maker (or controller);

(c) a family $\{A(x), x \in S\}$ of nonempty subsets $A(x)$ of A, where $A(x)$ denotes the set of actions available to a controller when the system is in state $x \in S$, and it is assumed to be Borel-measurable, that is, $A(x) \in \mathcal{B}(A)$ for every $x \in S$;

(d) a Borel measurable function $c(x,a)$ on K, called the cost rates, where $K := \{(x,a)|x \in S, a \in A(x)\}$ denotes the set of all feasible state-action pairs and is assumed in $\mathcal{B}(S \times A)$; (*As* $c(x,a)$ *is allowed to take positive and negative values, it can be interpreted as rewards rather than "costs" only.*)

(e) transition rates $q(dy|x,a)$, a universally measurable signed kernel on S given K. That is, $q(\cdot|x,a)$ satisfies countable additivity; $q(D|x,a) \geq 0$ for all $D \in \mathcal{B}(S)$ with $(x,a) \in K$ and $x \notin D$, being conservative in the sense of $q(S|x,a) \equiv 0$, and stable in that of

$$q^*(x) := \sup_{a\in A(x)} q(x,a) < \infty \quad \forall\, x \in S, \tag{2.2}$$

where $q(x,a) := -q(\{x\}|x,a) \geq 0$ for all $(x,a) \in K$.

Next, we give an informal description of the evolution of CTMDPs with model (2.1).

Roughly speaking, CTMDPs evolve as follows: A controller observes states of a system continuously in time. If the system is at state x_t at time t, he/she chooses an action

$a_t \in A(x_t)$ according to a given policy, as a consequence of which, the following happen:

(i) An immediate cost takes place at the rate $c(x_t, a_t)$.

(ii) After a random sojourn time (i.e., the holding time at state x_t), the system jumps to a set D ($x_t \notin D$) of states with the transition probability $\frac{q(D|x_t,a_t)}{q(x_t,a_t)}$ determined by the transition rates $q(dy|x_t, a_t)$. The distribution function of the sojourn time is $(1 - e^{-\int_t^{t+x} q(x_s,a_s)ds})$.

To formalize what is described above, below we describe the construction of CTMDPs under possibly randomized history-dependent policies.

To construct the underlying CTMDPs, we introduce some notation: Let $\Omega_0 := (S \times (0,\infty))^\infty$, $\Omega_k := (S \times (0,\infty))^k \times S \times (\{\infty\} \times \{\Delta\})^\infty$ for $k \geq 1$ and some $\Delta \notin S$, $\Omega := \cup_{k=0}^\infty \Omega_k$. $\mathcal{F}$ is the Borel σ-algebra on the Borel space Ω. Then we obtain the measurable space $(\Omega, \mathcal{F})$. For some $k \geq 1$, and sample $\omega := (x_0, \theta_1, x_1, \ldots, \theta_k, x_k, \infty, \Delta) \in \Omega$, define

$$T_k(\omega) := \theta_1 + \theta_2 + \ldots + \theta_k,\ T_\infty(\omega) := \lim_{k\to\infty} T_k(\omega),\ \text{and } X_k(\omega) := x_k. \tag{2.3}$$

In what follows, the argument ω is always omitted except some special informational statements. Then, we define the state process $\{x_t(\omega), t \geq 0\}$ on $(\Omega, \mathcal{F})$ by

$$x_t(\omega) := \sum_{k\geq 0} I_{\{T_k \leq t < T_{k+1}\}} X_k(\omega) + I_{\{t \geq T_\infty\}} \Delta, \quad \text{for } t \geq 0, \ \ (\text{with } T_0 := 0). \tag{2.4}$$

Obviously, $x_t(\omega)$ is right-continuous on $[0, \infty)$. We denote $x_{t-}(\omega) := \lim_{s\to t-} x_s(\omega)$. Here we have used the convenience that $0z =: 0$ and $0 + z =: z$ for all $z \in S_\Delta := S \cup \{\Delta\}$.

For each fixed $\omega := (x_0, \theta_1, x_1, \ldots, \theta_k, x_k, \ldots) \in \Omega$, from Eq. 2.4, we see that $T_k(\omega)$ ($k \geq 1$) denotes the k-th jump moment of $\{x_t, t \geq 0\}$, $X_{k-1}(\omega) = x_{k-1}$ is the state of the process on $[T_{k-1}(\omega), T_k(\omega))$, $\theta_k = T_k(\omega) - T_{k-1}(\omega)$ plays the role of sojourn time at state x_{k-1}, and the sample path $\{x_t(\omega), t \geq 0\}$ has at most denumerable states $x_k (k = 0, 1, \ldots)$. We do not intend to consider the controlled process $\{x_t, t \geq 0\}$ after moment T_∞, and thus view it to be absorbed in the cemetery state Δ. Hence, we write $A_\Delta := A \cup \{a_\Delta\}$, $A(\Delta) := \{a_\Delta\}$, $q(\cdot|\Delta, a_\Delta) :\equiv 0$, $c(\Delta, a_\Delta) :\equiv 0$, where a_Δ is an isolated point.

To precisely define the optimality criterion, we need to introduce the concept of a policy in Guo et al. (2012), Guo and Piunovskiy (2011), Kitaev and Rykov (1995), and Piunovskiy and Zhang (2011). To do so, we recall some notation. Take the right-continuous family of σ-algebras $\{\mathcal{F}_t\}_{t\geq 0}$ with $\mathcal{F}_t := \sigma(\{T_k \leq s, X_k \in D\} : D \in \mathcal{B}(S), s \leq t, k \geq 0)$. As in Guo and Song (2011), Guo and Piunovskiy (2011), Kitaev and Rykov (1995), and Piunovskiy and Zhang (2011), let $\mathcal{P}$ be the σ-algebra of predictable sets on $\Omega \times [0, \infty)$ related to $\{\mathcal{F}_t\}_{t\geq 0}$, that is, $\mathcal{P} := \sigma(B \times [0, \infty), C \times (s, \infty) : B \in \mathcal{F}_0, C \in \mathcal{F}_{s-}, s > 0)$ with $\mathcal{F}_{s-} := \bigvee_{t<s} \mathcal{F}_t := \sigma(\mathcal{F}_t, t < s)$. A real-valued function on $\Omega \times [0, \infty)$ is called *predictable* if it is measurable with respect to $\mathcal{P}$.

Definition 2.1 A transition probability $\pi(da|\omega, t)$ from $(\Omega \times [0, \infty), \mathcal{P})$ onto $(A_\Delta, \mathcal{B}(A_\Delta))$ such that $\pi(A(x_{t-}(\omega))|\omega, t) \equiv 1$ is called a (*randomized history-dependent*) policy. A policy $\pi(da|\omega, t)$ is called randomized Markovian if $\pi(da|\omega, t) \equiv \pi(da|x_{t-}(\omega), t)$. We will denote such a Markov policy by $\pi_t(da|\cdot)$ for informational implication. A randomized Markov policy $\pi_t(da|\cdot)$ is called deterministic Markovian whenever there exists a A-valued universally measurable function $f(t, x)$ on $[0, \infty) \times S$ such that $\pi_t(da|x)$ is the Dirac measure at point $f(t, x) \in A(x)$ for all $t \geq 0$ and $x \in S$. Such a Markov policy will be denoted by f for simplicity.

We denote by Π the set of all randomized history-dependent policies, by Π_m^r the set of all randomized Markov policies, and by Π_m^d the set of all (deterministic) Markov policies.

For any initial distribution γ on S and policy $\pi \in \Pi$, as showing in Guo et al. (2012), Guo and Song (2011), Guo and Piunovskiy (2011), Kitaev and Rykov (1995), and Piunovskiy and Zhang (2011) but using the extension of the Ionescu Tulcea theorem (e.g., Proposition 7.45 in Bertsekas and Shreve 1996); we see that there exists a unique probability measure $\mathbb{P}_\gamma^\pi$ (depending on γ and π) on $(\Omega, \mathcal{F})$. Let $\mathbb{E}_\gamma^\pi$ be the corresponding expectation operator. In particular, $\mathbb{E}_\gamma^\pi$ and $\mathbb{P}_\gamma^\pi$ will be respectively written as $\mathbb{E}_x^\pi$ and $\mathbb{P}_x^\pi$ when γ is the Dirac measure at a state x in S.

Fix any finite horizon $T > 0$. For each policy $\pi \in \Pi$ and state $x \in S$, we define the T-horizon risk-sensitive criterion $J(\pi, 0, x)$ of CTMDPs by

$$J(\pi, 0, x) := \mathbb{E}_x^\pi \left[e^{\delta \int_0^T \int_A c(x_t, a)\pi(da|\omega, t)dt} \right], \tag{2.5}$$

provided that the integral is well defined, where δ is a constant called the risk-sensitive parameter. In the following arguments, we assume that $\delta > 0$. For the other case of $\delta < 0$, the corresponding arguments are similar, and thus omitted.

Note that the process $\{x_t, t \geq 0\}$ on $(\Omega, \mathcal{F}, \mathbb{P}_\gamma^\pi)$ may *not* be Markovian since the policy π can depend on histories $(x_0, \theta_1, x_1, \ldots, \theta_k, x_k)$. However, for each $\pi := \pi_t(da|\cdot) \in \Pi_m^r$, it is well known (e.g. Feinberg et al. 2014) that $\{x_t, t \geq 0\}$ is a Markov process on $(\Omega, \mathcal{F}, \mathbb{P}_\gamma^\pi)$, and thus for each $x \in S$ and $t \in [0, T]$, the following expression

$$J(\pi, t, x) := \mathbb{E}_\gamma^\pi \left[e^{\delta \int_t^T \int_A c(x_s, a)\pi_s(da|x_s)ds} | x_t = x \right],$$

is well defined (when the integral exists), and it is called the risk-sensitive value of π from the horizon t to T.

For each $x \in S$, let

$$J_*(t, x) := \inf_{\pi \in \Pi_m^r} J(\pi, t, x) \qquad \text{for } t \in [0, T]. \tag{2.6}$$

The function $J_*(t, x)$ on $[0, T] \times S$ is called the value function of the CTMDPs with the T-horizon risk-sensitive criterion.

Definition 2.2 A policy $\pi^* \in \Pi$ is said to be optimal if $J(\pi^*, 0, x) \leq J(\pi, 0, x)$ for all $\pi \in \Pi$ and $x \in S$.

The main goal of this paper is to give conditions for the existence of optimal Markov policies.

At this end, we give some remarks about the difference between the T-horizon risk-sensitive criterion $J(\pi, 0, x)$ and the risk-neutral one in Guo et al. (2015b) and Yushkevich (1977), which is defined by

$$V(\pi, 0, x) := \mathbb{E}_x^\pi \left[\int_0^T \int_A c(x_t, a)\pi(da|\omega, t)dt \right] =: \mathbb{E}_x^\pi Y, \text{ with } Y := \int_0^T \int_A c(x_t, a)\pi(da|\omega, t)dt,$$

where $c(x, a)$ is assumed to be bounded positive on K. Then, by the Jensen's inequality and the monotonicity of the log function, we have $\ln J(\pi, 0, x) \geq \delta V(\pi, 0, x)$. More, by

Taylor's expansion of $e^{\delta Y}$ at the point $z := \mathbb{E}_x^\pi Y$, we have

$$\begin{aligned} e^{\delta Y} &= e^{\delta z} + \delta e^{\delta z}(Y-z) + \frac{1}{2}\delta^2 e^{\delta z}(Y-z)^2 \\ &\quad + \frac{1}{6}\delta^3 e^{\delta \xi}(Y-z)^3 \text{ (with } \xi \text{ being between } Y \text{ and } z), \end{aligned}$$

which also implies that ξ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P}_\gamma^\pi)$. Therefore, we have

$$\begin{aligned} \ln J(\pi,0,x) &= \ln\left[e^{\delta z} + \frac{1}{2}e^{\delta z}\delta^2 var(Y) + \frac{1}{6}\delta^3 \mathbb{E}_x^\pi e^{\delta\xi}(Y-z)^3\right] \\ &= \delta V(\pi,0,x) + \ln\left[1 + \frac{1}{2}\delta^2 var(Y) + \frac{1}{6}\delta^3 \mathbb{E}_x^\pi e^{\delta(\xi-z)}(Y-z)^3\right] \\ &= \delta V(\pi,0,x) + \ln[1 + \frac{1}{2}\delta^2 var(Y)] + o(\delta^2), \end{aligned} \tag{2.7}$$

where $var(Y) := \mathbb{E}_x^\pi(Y - \mathbb{E}_x^\pi Y)^2$ denoting the variance of Y, and the corresponding term $\ln[1 + \frac{1}{2}\delta^2 var(Y)]$ is called a risk premium. Thus, the difference between $\ln J(\pi, 0, x)$ and $\delta V(\pi, 0, x)$ is the summation of the risk premium plus the other small order term. Therefore, the risk-sensitive criterion is more risk-sensitive than the risk-neutral one because of its inclusion of the risk premium.

From Eq. 2.7 we see that $\lim_{\delta\to 0} \frac{\ln J(\pi,0,x)}{\delta} = V(\pi, 0, x)$. On the other hand, for any fixed $\delta > 0$, since $\inf_{\pi\in\Pi} \ln J(\pi, 0, x) = \ln J(\pi^*, 0, x)$ *if and only if* $\inf_{\pi\in\Pi} J(\pi, 0, x) = J(\pi^*, 0, x)$, we will consider $J(\pi, 0, x)$, instead of $\ln J(\pi, 0, x)$.

3 Preliminaries

This section provides some preliminary facts for our arguments below.

Since the rates $q(dy|x, a)$ and costs $c(x, a)$ are allowed to be unbounded, we next give conditions for the non-explosion of $\{x_t, t \geq 0\}$ and finiteness of $J(\pi, t, x)$.

Assumption 3.1 There exist a real-valued Borel measurable function $V_0 \geq 1$ on S and constants $\rho_0, b_0 \geq 0$, $M_0', M_0 \geq 1$, such that

(i) $\int_S V_0(y)q(dy|x, a) \leq \rho_0 V_0(x) + b_0$, for all $(x, a) \in K$;
(ii) $q^*(x) \leq M_0' V_0(x)$ for all $x \in S$, where $q^*(x)$ is as in Eq. 2.2;
(iii) $e^{2T\delta|c(x,a)|} \leq M_0 V_0(x)$ for all $(x, a) \in K$, with T and δ as in Eq. 2.5.

Remark 3.1 (a) Assumptions 3.1(i,ii) are used and verified with examples in Guo and Hernández-Lerma (2009), Guo and Piunovskiy (2011), Piunovskiy and Zhang (2011), and Prieto-Rumeau and Hernández-Lerma (2012) for the risk-neutral CTMDPs. When the transition rates are bounded (i.e., $\|q^*\| < \infty$) (Ghosh and Saha 2014; Kitaev and Rykov 1995; Kumar and Chandan 2013, 2015; Yushkevich 1977), Assumptions 3.1(i-ii) are satisfied by taking $V_0(x) \equiv 1$.

(b) Assumption 3.1(iii) is for the finiteness of the value function $J_*(t, x)$, and it is new and satisfied when $c(x, a)$ is bounded (i.e., $\|c\| := \sup_{(x,a)\in K} |c(x, a)| < \infty$) (Ghosh and Saha 2014; Kumar and Chandan 2013, 2015; Wei 2016; Yushkevich 1977). Moreover, if $\delta|c(x, a)| \leq \sqrt{\ln V_0(x)} + L$ for all $(x, a) \in K$ and some constant $L \geq 0$, then Assumption 3.1(iii) holds: Indeed, since $\delta|c(x, a)| \leq \sqrt{\ln V_0(x)} + L \leq$

$\frac{t}{2}+\frac{\ln V_0(x)}{2t}+L$ for all $(x,a)\in K$ and $t>0$, we have $e^{2T\delta|c(x,a)|}\le e^{T^2+\ln V_0(x)+2TL}=e^{T^2+2TL}V_0(x)$, which implies Assumption 3.1(iii) with $M_0:=e^{T^2+2TL}$.

(c) If the number ρ_0 in Assumption 3.1(i) is not positive, then Assumption 3.1(i) still holds when ρ_0 is replaced with the positive number "$1+|\rho_0|$". Thus, it is just for convenience to assume that the constant $\rho_0>0$ throughout the following. However, the corresponding number is assumed to be negative in Guo and Hernández-Lerma (2009), Guo et al. (2012), and Prieto-Rumeau and Hernández-Lerma (2012) or less than the discount factor in Guo and Hernández-Lerma (2009) and Guo and Piunovskiy (2011).

Lemma 3.1 *Under Assumption 3.1, for each $\pi\in\Pi$, the following assertions hold.*

(a) $\mathbb{P}^\pi_x(x_t\in S)=1$, $\mathbb{P}^\pi_x(T_\infty=\infty)=1$, *and* $\mathbb{P}^\pi_x(x_0=x)=1$ *for each* $t\ge 0$ *and* $x\in S$;

(b) $\mathbb{E}^\pi_x[V_0(x_t)]\le e^{\rho_0 t}[V_0(x)+\frac{b_0}{\rho_0}]$, *for each* $t\ge 0$, $x\in S$ *and* $\pi\in\Pi$;

(c) $\mathbb{E}^\pi_\gamma[V_0(x_t)|x_s=x]\le e^{\rho_0(t-s)}[V_0(x)+\frac{b_0}{\rho_0}]$, *for each* $t\ge s\ge 0$, $x\in S$ *and* $\pi\in\Pi^r_m$.

(d) *If, in addition, Assumption 3.1 (iii) is satisfied, then*

(d_1) $e^{-LV_0(x)}\le J(\pi,0,x)\le LV_0(x)$ *for* $x\in S$ *and* $\pi\in\Pi$, *where* $L:=M_0e^{\rho_0 T}[1+\frac{b_0}{\rho_0}]$;

(d_2) $e^{-LV_0(x)}\le J(\pi,t,x)\le LV_0(x)$ *for* $(t,x)\in[0,T]\times S$ *and* $\pi\in\Pi^r_m$.

Proof Parts (a) and (b) follow from Guo et al. (2012) and Guo and Piunovskiy (2011) (or Piunovskiy and Zhang 2011); while part (c) is from Theorem 3.1 in Guo (2007). We next prove part (d). For almost surely (a.s.) $\omega:=(x_0,\theta_1,x_1,\ldots,\theta_k,x_k,\ldots)\in\Omega$ (with respect to $\mathbb{P}^\pi_x$), let k (depending on ω) be determined by $T_k(\omega)\le T<T_{k+1}(\omega)$. Then, it follows from (a) and Eq. 2.4 that k is finite, and $\{x_t(\omega), t\in[0,T]\}=\{X_0(\omega),\ldots,X_k(\omega)\}$. Thus, since $|c(x,a)|\le\frac{1}{T\delta}\ln\sqrt{M_0V_0(x)}$ (by Assumption 3.1(iii)) and $\theta_{i+1}=T_{i+1}(\omega)-T_i(\omega)(i=0,\ldots,k-1)$, we have

$$\begin{aligned}&\int_0^T\int_A|c(x_t,a)|\pi(da|\omega,t)dt\\ &\le\frac{\sum_{i=0}^{k-1}\theta_{i+1}\ln\sqrt{M_0V_0(X_i(\omega))}+(T-T_k(\omega))\ln\sqrt{M_0V_0(X_k(\omega))}}{T\delta},\end{aligned}$$

which implies that $\int_0^s\int_A c(x_t,a)\pi(da|\omega,t)dt$ is real-valued Borel measurable in $s\in[0,T]$ (for the given ω). Thus, by the Jensen inequality with respect to the probability measure $\frac{dt}{T}$ on $\mathcal{B}([0,T])$, we have

$$e^{\int_0^T\delta\int_A c(x_t,a)\pi(da|\omega,t)dt}\le\frac{1}{T}\int_0^T e^{T\delta\int_A|c(x_t,a)|\pi(da|\omega,t)}dt,\quad a.s.-\mathbb{P}^\pi_x.$$

Therefore (by (b)),

$$\begin{aligned}\mathbb{E}^\pi_x\left[e^{\int_0^T\delta\int_A c(x_t,a)\pi(da|\omega,t)dt}\right]&\le\mathbb{E}^\pi_x\left[\frac{1}{T}\int_0^T e^{T\delta\int_A|c(x_t,a)|\pi(da|\omega,t)}dt\right]\\ &\le\frac{1}{T}M_0\int_0^T\mathbb{E}^\pi_x\left[V_0(x_t)\right]dt\le M_0e^{\rho_0 T}[V_0(x)+\frac{b_0}{\rho_0}].\end{aligned}$$

Hence, by Eq. 2.5, we have

$$J(\pi,0,x) \le M_0 e^{\rho_0 T}\left[1+\frac{b_0}{\rho_0 V_0(x)}\right]V_0(x) \le LV_0(x). \tag{3.1}$$

On the other hand, since $T\delta|c(x,a)| \le e^{2T\delta|c(x,a)|} \le M_0V_0(x)$, we have

$$\begin{aligned} J(\pi,0,x) = \ & \mathbb{E}_x^\pi\left[e^{\delta\int_0^T\int_A c(x_t,a)\pi(da|\omega,t)dt}\right] \ge e^{\delta\mathbb{E}_x^\pi\left[\int_0^T\int_A c(x_t,a)\pi(da|\omega,t)dt\right]} \\ & \ge e^{-\frac{1}{T}\mathbb{E}_x^\pi\left[\int_0^T\int_A M_0V_0(x_t)\pi(da|\omega,t)dt\right]} = e^{-\frac{1}{T}\mathbb{E}_x^\pi\left[\int_0^T M_0V_0(x_t)dt\right]} \ge e^{-\frac{M_0}{T}\int_0^T\mathbb{E}_x^\pi[V_0(x_t)]dt} \\ & \ge e^{-\frac{M_0}{T}\int_0^T e^{\rho_0 T}[V_0(x)+\frac{b_0}{\rho_0}]dt} = e^{-M_0e^{\rho_0 T}[V_0(x)+\frac{b_0}{\rho_0}]} \ge e^{-LV_0(x)}, \end{aligned}$$

which, together with Eq. 3.1, implies (d$_1$). Similarly, we see that (d$_2$) is also true. □

Lemma 3.1 gives conditions for the finiteness of $J(\pi,t,x)$ as well as the non-explosion of $\{x_t, t \ge 0\}$. In order to deal with the optimality for history-dependent policies, we need the following facts, which extend both the *analog* of Ito-Dynkin's formula in Guo et al. (2015b) and Feynman-Kac formula in Wei (2016) to a more general case of possible non-Markov processes $\{x_t, t \ge 0\}$ and functions $\varphi(\omega,t,x)$ with an additional element $\omega \in \Omega$.

Denote by m_L the Lebesgue's measure on $[0,T]$, and by $\mathbb{B}_{\mathcal{P}}(\Omega\times[0,T]\times S)$ the set of real-valued and $\mathcal{P}\times\mathcal{B}(S)$-measurable functions $\varphi(\omega,t,x)$ with the following features: Given any $x\in S, \pi\in\Pi$, and a.s. $\omega\in\Omega$ with respect to $\mathbb{P}_x^\pi$, there exists a Borel subset $E_{(\varphi,\omega,x,\pi)}$ (depending on the φ,ω,x,π) of $[0,T]$ such that the partial derivative $\varphi'(\omega,t,x)$ (with respect to t) exists for every $t\in E_{(\varphi,\omega,x,\pi)}$ and $m_L(E^c_{(\varphi,\omega,x,\pi)})=0$. Obviously, if a function $\varphi(\omega,t,x)$ in $\mathbb{B}_{\mathcal{P}}(\Omega\times[0,T]\times S)$ is independent of ω (written as $\varphi(t,x)$), and so is the corresponding $E_{(\varphi,\omega,x,\pi)}$, which will be denoted by $E_{(\varphi,x)}$.

For any $\varphi(\omega,t,x)\in\mathbb{B}_{\mathcal{P}}(\Omega\times[0,T]\times S)$, when the partial derivative does not exist for some $(\omega,t,x)\in\Omega\times[0,T]\times S$, for simplicity of arguments, we take $\varphi'(\omega,t,x)$ to be any real number, and so φ' is defined on $\Omega\times[0,T]\times S$. We will see that such a modification of φ' loses nothing. For example, $\mathbb{E}_x^\pi\left[\int_0^T|\varphi'(\omega,t,x_t)|dt\right]$ and $\int_0^T|\varphi'(\omega,t,x_t)|dt$ are defined well for each $(x,\pi)\in S\times\Pi$ and a.s. $\omega\in\Omega$ with respect to $\mathbb{P}_x^\pi$, respectively. Indeed, for each $x\in S, \pi\in\Pi$, and $s\in[0,T]$, Lemma 3.1(a) together with Eq. 2.4 gives the existence of a Boreal measurable subset $\hat\Omega$ of Ω such that: 1) $\mathbb{P}_x^\pi(\hat\Omega^c)=0$, and 2) for each given $\omega\in\hat\Omega$, $x_t(\omega)=x_i$ for $T_i(\omega)\le t<T_{i+1}(\omega)(0\le i\le k)$ for some a finite k, where k and $x_i\in S$ depend on ω, and k is determined by $T_k(\omega)\le T<T_{k+1}(\omega)$. Let $\tilde i := \min\{l : s < T_l\}$ (depending on ω and s). Then, since $m_L(\cup_{i=0}^k E^c_{(\varphi,\omega,x_i,\pi)})=0$, for each $\omega\in\hat\Omega$,

$$\int_s^T|\varphi'(\omega,t,x_t)|dt = \int_s^{T_{\tilde i}}|\varphi'(\omega,t,x_{\tilde i-1})|dt + \sum_{l=\tilde i}^{k-1}\int_{T_l}^{T_{l+1}}|\varphi'(\omega,t,x_l)|dt + \int_{T_k}^T|\varphi'(\omega,t,x_k)|dt$$

is well defined. Therefore (by $\mathbb{P}_x^\pi(\hat\Omega^c)=0$), $\mathbb{E}_x^\pi\left[\int_0^T|\varphi'(\omega,t,x_t)|dt\right]$ is also well defined.

Lemma 3.2 *Under Assumption 3.1, the following assertions hold.*

(a) *For any given $\pi \in \Pi$, $x \in S$, if a function $\varphi \in \mathbb{B}_{\mathcal{P}}(\Omega \times [0,T] \times S)$ satisfies that*

$$\mathbb{E}_x^\pi\left[\int_0^T \int_A \int_S |\varphi(\omega,t,y)||q|(dy|x_t,a)\pi(da|\omega,t)dt\right] \\ +\mathbb{E}_x^\pi\left[\int_0^T |\varphi'(\omega,t,x_t)|dt\right] < \infty, \tag{3.2}$$

where $|q|(dy|x,a) := \int q(dy\backslash\{x\})|x,a) - q(\{x\}|x,a)\delta_x(dy)$ for all $(x,a) \in K$, then

$$\mathbb{E}_x^\pi\left[\int_0^T \left(\varphi'(\omega,t,x_t) + \int_A \int_S \varphi(\omega,t,y)q(dy|x_t,a)\pi(da|\omega,t)\right)dt\right] \\ = \mathbb{E}_x^\pi \varphi(\omega,T,x_T) - \mathbb{E}_x^\pi \varphi(\omega,0,x).$$

(b) *For any $\pi = \pi_t(da|\cdot) \in \Pi_m^r$ and a Borel measurable function $\phi \in \mathbb{B}_{\mathcal{P}}([0,T] \times S)$, such that the corresponding function $\varphi(\omega,t,x) := e^{\int_s^t \delta c(x_v(\omega),\pi_v)dv}\phi(t,x)$ satisfying (3.2), then*

$$\mathbb{E}_\gamma^\pi\left[\int_s^T \left(\left(e^{\int_s^t \delta c(x_v,\pi_v)dv}\phi(t,x_t)\right)' + \int_S e^{\int_s^t \delta c(x_v,\pi_v)dv}\phi(t,y)q(dy|x_t,\pi_t)\right)dt|x_s = x\right]$$

$$= \mathbb{E}_\gamma^\pi\left[e^{\int_s^T \int_A \delta c(x_v,a)\pi_v(da|x_v)dv}\phi(T,x_T)|x_s = x\right] - \phi(s,x),$$

where $c(x,\pi_t) := \int_{A(x)} c(x,a)\pi_t(da|x)$ and $q(dy|x,\pi_t) := \int_{A(x)} q(dy|x,a)\pi_t(da|x)$.

(c) *If the functions $c(x,a)$ and $\phi(t,x)$ in (b) above are both bounded and $|\phi'(t,x)| \le CV_0(x)$ for all $(t,x) \in [0,T] \times S$ with some constant $C > 0$, then*

$$\mathbb{E}_\gamma^\pi\left[\int_s^T \left(\left(e^{\int_s^t \delta c(x_v,\pi_v)dv}\phi(t,x_t)\right)' + \int_S e^{\int_s^t \delta c(x_v,\pi_v)dv}\phi(t,y)q(dy|x_t,\pi_t)\right)dt|x_s = x\right]$$

$$= \begin{cases} \mathbb{E}_x^\pi\left(e^{\int_s^T \int_A \delta c(x_v,a)\pi_v(da|x_v)dv}\phi(T,x_T)\right) - \phi(0,x), & \text{for } s = 0, \pi \in \Pi \\ \mathbb{E}_\gamma^\pi\left[e^{\int_s^T \int_A \delta c(x_v,a)\pi_v(da|x_v)dv}\phi(T,x_T)|x_s = x\right] - \phi(s,x), & \text{for } s \in [0,T], \pi \in \Pi_m^r. \end{cases}$$

Proof (a) Lemma 3.1(a) together with Eq. 2.4 gives the existence of a Boreal measurable subset $\hat{\Omega}$ of Ω such that $\mathbb{P}_x^\pi(\hat{\Omega}^c) = 0$. For each $\omega \in \hat{\Omega}$, for $0 \le t \le T$, from definitions (2.3) and (2.4), denote $k_t(\omega) := \max\{k|T_k(\omega) \le t\}$. Since $x_t(\omega)$ is right-continuous on $[0,\infty)$, we have $x_t = x_{T_{k_t}}$ when $T_{k_t} \le t < T_{k_t+1}$. Since $\int_s^T \varphi'(\omega,t,x_t)dt$ is well defined (just proved), for $0 \le s \le T$ and $\omega \in \hat{\Omega}$, we have

$$\varphi(\omega,T,x_T) = \varphi(\omega,s,x_s) + \int_s^T \varphi'(\omega,t,x_t)dt + \sum_{n=k_s+1}^{k_T} \int_{(s,T]} \Delta\varphi(\omega,t,x_t)\delta_{T_n}(dt) \quad \mathbb{P}_x^\pi - a.s. \tag{3.3}$$

Denote $\Delta\varphi(\omega, t, x_t) := \varphi(\omega, t, x_t(\omega)) - \varphi(\omega, t, x_{t-}(\omega))$, and recall that $x_{t-}(\omega) = \lim_{s\to t-} x_s(\omega)$. Equation 3.3 is proved as follows.

$$\begin{aligned}
&\varphi(\omega, T, x_T)\\
=\ &\varphi(\omega, s, x_s) + \varphi(\omega, T_{k_s+1}, x_{T_{k_s+1}-}) - \varphi(\omega, s, x_s)\\
&+ \sum_{n=k_s+1}^{k_T} \big[\varphi(\omega, T_n, x_{T_n}) - \varphi(\omega, T_n, x_{T_n-})\big] + \sum_{n=k_s+1}^{k_T-1} \big[\varphi(\omega, T_{n+1}, x_{T_{n+1}-}) - \varphi(\omega, T_n, x_{T_n})\big]\\
&+\varphi(\omega, T, x_{T-}) - \varphi(\omega, T_{k_T}, x_{T_{k_T}}) + \varphi(\omega, T, x_T) - \varphi(\omega, T, x_{T-})\\
=\ &\varphi(\omega, s, x_s) + \varphi(\omega, T_{k_s+1}, x_s) - \varphi(\omega, s, x_s)\\
&+ \sum_{n=k_s+1}^{k_T} \big[\varphi(\omega, T_n, x_{T_n}) - \varphi(\omega, T_n, x_{T_n-})\big] + \sum_{n=k_s+1}^{k_T-1} \big[\varphi(\omega, T_{n+1}, x_{T_n}) - \varphi(\omega, T_n, x_{T_n})\big]\\
&+\varphi(\omega, T, x_{T_{k_T}}) - \varphi(\omega, T_{k_T}, x_{T_{k_T}}) + \varphi(\omega, T, x_T) - \varphi(\omega, T, x_{T-})\\
=\ &\varphi(\omega, s, x_s) + \varphi(\omega, T_{k_s+1}, x_s) - \varphi(\omega, s, x_s) + \sum_{n=k_s+1}^{k_T-1} \big[\varphi(\omega, T_{n+1}, x_{T_n}) - \varphi(\omega, T_n, x_{T_n})\big]\\
&+\varphi(\omega, T, x_{T_{k_T}}) - \varphi(\omega, T_{k_T}, x_{T_{k_T}})\\
&+ \sum_{n=k_s+1}^{k_T} \big[\varphi(\omega, T_n, x_{T_n}) - \varphi(\omega, T_n, x_{T_n-})\big] + \varphi(\omega, T, x_T) - \varphi(\omega, T, x_{T-})\\
=\ &\varphi(\omega, s, x_s) + \int_{[s, T_{k_s+1})} \varphi'(\omega, t, x_s)dt + \sum_{n=k_s+1}^{k_T-1} \int_{[T_n, T_{n+1})} \varphi'(\omega, t, x_{T_n})dt\\
&+ \int_{[T_{k_T}, T)} \varphi'(\omega, t, x_{T_{k_T}})dt + \sum_{n=k_s+1}^{k_T} \Delta\varphi(\omega, T_n, x_{T_n}) + \Delta\varphi(\omega, T, x_T)\\
=\ &\varphi(\omega, s, x_s) + \int_{[s, T_{k_s+1})} \varphi'(\omega, t, x_t)dt + \sum_{n=k_s+1}^{k_T-1} \int_{[T_n, T_{n+1})} \varphi'(\omega, t, x_t)dt\\
&+ \int_{[T_{k_T}, T)} \varphi'(\omega, t, x_t)dt + \sum_{n=k_s+1}^{k_T} \int_{(s,T]} \Delta\varphi(\omega, t, x_t)\delta_{T_n}(dt)\\
=\ &\varphi(\omega, s, x_s) + \int_s^T \varphi'(\omega, t, x_t)dt + \sum_{n=k_s+1}^{k_T} \int_{(s,T]} \Delta\varphi(\omega, t, x_t)\delta_{T_n}(dt) \quad \mathbb{P}_x^\pi - a.s.
\end{aligned}$$

Then, Lemma 4.28 in Kitaev and Rykov (1995) shows that the random measure m^π defined by

$$m^\pi(B|\omega, t)dt := \int_A q(B|x_{t-}, a)\pi(da|\omega, t)I_{\{x_{t-}\notin B\}}dt, \qquad B \in \mathcal{B}(S) \tag{3.4}$$

is the dual predictable projection of the random measure $\sum_{n\geq 1} \delta_{(T_n, X_{n-1})}(dt, dx)$ on $\mathcal{B}((0,\infty)\times S)$ under $\mathbb{P}_x^\pi$. Thus, by the definition of a dual predictable projection in (4.5) in Kitaev and Rykov (1995), we have

$$\begin{aligned}
&\mathbb{E}_x^\pi\left[\sum_{n\geq 1}\int_{(0,T]} \Delta\varphi(\omega, t, x_t)\delta_{T_n}(dt)\right]\\
=\ &\mathbb{E}_x^\pi\left[\int_S\int_{(0,T]} (\varphi(\omega, s, y) - \varphi(\omega, s, x_{s-}))m^\pi(dy|\omega, s)ds\right]
\end{aligned}$$

which, together with Eq. 3.4 and the expectation of both sides of Eq. 3.3 with $s = 0$, gives

$$\begin{aligned}
&\mathbb{E}_x^\pi\left[\varphi(\omega, T, x_T)\right]\\
&= \mathbb{E}_x^\pi \varphi(\omega, 0, x) + \mathbb{E}_x^\pi\left[\int_0^T \varphi'(\omega, t, x_t)dt\right] + \mathbb{E}_x^\pi\left[\sum_{n\geq 1}\int_{(0,T]} \Delta\varphi(\omega, t, x_t)\delta_{T_n}(dt)\right]\\
&= \mathbb{E}_x^\pi \varphi(\omega, 0, x) + \mathbb{E}_x^\pi\left[\int_0^T \varphi'(\omega, t, x_t)dt\right] + \mathbb{E}_x^\pi\left[\int_{(0,T]}\int_S [\varphi(\omega, s, y) - \varphi(\omega, s, x_{s-})]m^\pi(dy|\omega, s)ds\right]\\
&= \mathbb{E}_x^\pi \varphi(\omega, 0, x) + \mathbb{E}_x^\pi\left[\int_0^T \varphi'(\omega, t, x_t)dt\right] + \mathbb{E}_x^\pi\left[\int_{(0,T]}\int_S\int_A \varphi(\omega, t, y)q(dy|x_{t-}, a)\pi(da|\omega, t)dt\right].
\end{aligned}$$

Here, integrability results such as Eq. 3.2 validate all the involved operations. Moreover, since $x_{t-}(\omega) = x_t(\omega)$ on $(0, T]$ except finite time points t, part (a) follows.

(b) In lieu of $\varphi(\omega, t, x_t)$ with $e^{\int_s^t \int_A \delta c(v,x_v,a)\pi_v(da|x_v)dv}\phi(t, x_t)$, taking the conditional expectation $\mathbb{E}_\gamma^\pi[\cdot|x_s = x]$ in both sides of Eq. 3.3 and using the Markov property of $\{x_t, t \geq 0\}$, as the proof of (a), we see that (b) is also true.

(c) Under the condition in (c), the function $\varphi(\omega, t, x)$ in (b) is bounded on $\Omega \times [0, T] \times S$, and thus the first part of Eq. 3.2 is finite (by Lemma 3.1(b) and Assumptions 3.1(i,ii)). Moreover, since

$$\begin{aligned}
\left(e^{\int_s^t \int_A |\delta c(x_v,a)|\pi_v(da|x_v)dv}\phi(t, x_t)\right)' &= \delta|c(x_t, \pi_t)|e^{\int_s^t \int_A |\delta c(x_v,a)|\pi_v(da|x_v)dv}\phi(t, x_t)\\
&\quad + e^{\int_s^t \int_A |\delta c(x_v,a)|\pi_v(da|x_v)dv}\phi'(t, x_t)\\
&\leq \delta\|c\|e^{T\|c\|}\|\phi\| + Ce^{T\delta\|c\|}V_0(x_t)
\end{aligned}$$

which, together with Lemma 3.1(b), implies that the second part of Eq. 3.2 is finite. Thus, (c) follows.

□

Next we derive an extension of Dynkin's formula by Lemma 3.2. To do so, we introduce the following condition and notation.

Assumption 3.2 There exist a Borel measurable function $V_1 \geq 1$ on S, and positive constants ρ_1, b_1, and M_1, such that

(i) $\int_S V_1^2(y)q(dy|x, a) \leq \rho_1 V_1^2(x) + b_1$ for all $(x, a) \in K$;
(ii) $V_0^2(x) \leq M_1 V_1(x)$ for all $x \in S$, with the $V_0(x)$ as the Assumption 3.1(i).

Assumption 3.2 is used to give a domain for the Dynkin's formula below, and it is obviously satisfied when the transition rates are bounded (Ghosh and Saha 2014; Kumar and Chandan 2013, 2015; Yushkevich 1977).

Given the $V_k(k = 0, 1)$ as in Assumption 3.2 and any Borel set Z, a real-valued function φ on $Z \times S$ is called V_k-bounded if the V_k-weighted norm of φ, $\|\varphi\|_{V_k} := \sup_{(z,x)\in Z\times S} \frac{|\varphi(z,x)|}{V_k(x)}$, is finite. We denote by $\mathbb{B}_{V_k}(Z \times S)$ the Banach space of all V_k-bounded functions on $Z \times S$. When $V_k(x) \equiv 1$, $\mathbb{B}_1(Z \times S)$ is the space of all bounded functions. In particular, take $Z = \Omega \times [0, T]$ or $[0, T]$, we define

$$\mathbb{B}^1_{V_0,V_1}(\Omega \times [0, T] \times S) := \{\varphi \in \mathbb{B}_{V_0}(\Omega \times [0, T] \times S) \cap \mathbb{B}_{\mathcal{P}}(\Omega \times [0, T] \times S) \mid \varphi' \in \mathbb{B}_{V_1}(\Omega \times [0, T] \times S)\},$$

and then

$$\mathbb{B}^1_{V_0,V_1}([0,T]\times S) := \{\varphi \in \mathbb{B}_{V_0}([0,T]\times S) \cap \mathbb{B}_{\mathcal{P}}([0,T]\times S) \mid \varphi' \in \mathbb{B}_{V_1}([0,T]\times S)\}.$$

Theorem 3.1 below is an extension of Theorem 3.1 in Guo et al. (2015b) and Theorem 3.1 in Wei (2016) from the forms of functions $\varphi(t,x)$ to the more general forms of $\varphi(\omega,t,x)$ with the additional variable ω, and the extension is needed for the following arguments over history-dependent policies. The proof of Theorem 3.1 is based on Lemma 3.2, which is different from those in Guo et al. (2015b) and Wei (2016) for Markov policies only.

Theorem 3.1 *Suppose Assumptions 3.1 and 3.2 are satisfied. Then, for each $(s,x)\in[0,T]\times S$, the following assertions hold.*

(a) (The extension of Dynkin's formula): For every $\pi\in\Pi$ and $\varphi\in\mathbb{B}^1_{V_0,V_1}(\Omega\times[0,T]\times S)$,

$$\mathbb{E}^\pi_x\left[\int_0^T\left(\varphi'(\omega,t,x_t)+\int_A\int_S\varphi(\omega,t,y)q(dy|x_t,a)\pi(da|\omega,t)\right)dt\right]$$
$$=\mathbb{E}^\pi_x\varphi(\omega,T,x_T)-\mathbb{E}^\pi_x\varphi(\omega,0,x),$$

where $\{x_t, t\geq 0\}$ may be not Markovian since the policy π may depend on histories.

(b) (The Dynkin's formula): For each $\pi\in\Pi^r_m$, and $\varphi\in\mathbb{B}^1_{V_0,V_1}([0,T]\times S)$,

$$\mathbb{E}^\pi_\gamma\left[\int_s^T\left(\left(e^{\int_s^t\delta c(x_v,\pi_v)dv}\varphi(t,x_t)\right)'+\int_S e^{\int_s^t\delta c(x_v,\pi_v)dv}\varphi(t,y)q(dy|x_t,\pi_t)\right)dt|x_s=x\right]$$
$$=\mathbb{E}^\pi_\gamma\left[e^{\int_s^T\delta c(x_t,\pi_t)dt}\varphi(T,x_T)|x_s=x\right]-\varphi(s,x).$$

Proof (a) By the definition $\mathbb{B}^1_{V_0,V_1}(\Omega\times[0,T]\times S)$ and $\varphi\in\mathbb{B}^1_{V_0,V_1}(\Omega\times[0,T]\times S)$, we have

$$|\varphi(\omega,t,x)|\leq\|\varphi\|_{V_0}V_0(x),\ |\varphi'(\omega,t,z)|\leq\|\varphi'\|_{V_1}V_1(x)\quad\forall\,(\omega,t,x)\in\Omega\times[0,T]\times S. \tag{3.5}$$

Thus, by Assumptions 3.1(i,ii) and 3.2 we have, for $(\omega,t,z)\in\Omega\times[0,T]\times S$

$$\begin{aligned}\int_A\int_S|\varphi(\omega,t,y)||q|(dy|z,a)\pi(da|\omega,t) &\leq \|\varphi\|_{V_0}\left[\int_A\int_S V_0(y)|q|(dy|z,a)\pi(da|\omega,t)\right]\\ &\leq \|\varphi\|_{V_0}[\rho_0V_0(z)+b_0+2M_0'V_0^2(z)]\\ &\leq \|\varphi\|_{V_0}[\rho_0M_1+b_0+2M_0'M_1]V_1(z). \end{aligned}\tag{3.6}$$

Moreover, by Eq. 3.5 we have

$$\int_0^T|\varphi'(\omega,t,x_t)|dt\leq\|\varphi'\|_{V_1}\int_0^TV_1(x_t)dt\leq\|\varphi'\|_{V_1}\int_0^TV_1^2(x_t)dt,$$

which, together with Lemma 3.1(b), gives

$$\begin{aligned}\mathbb{E}^\pi_x\left[\int_0^T|\varphi'(\omega,t,x_t)|dt\right] &\leq \|\varphi'\|_{V_1}\mathbb{E}^\pi_x\left[\int_0^TV_1^2(x_t)dt\right]\\ &\leq \|\varphi'\|_{V_1}Te^{\rho_1T}\left[\frac{V_1^2(x)}{T\rho_1}+\frac{b_1}{\rho_1}\right]<\infty.\end{aligned}\tag{3.7}$$

Thus, by Eqs. 3.6–3.7 and Lemma 3.1(b) we have

$$\begin{aligned}&\mathbb{E}_x^\pi\left[\int_0^T\int_A\int_S|\varphi(\omega,t,y)||q|(dy|x_t,a)\pi(da|\omega,t)dt\right]\\&\leq T\|\varphi\|_{V_0}[\rho_0M_1+b_0+2M_0'M_1]e^{\rho_1T}\left[\frac{V_1^2(x)}{T\rho_1}+\frac{b_1}{\rho_1}\right],\end{aligned}$$

which, together with Eq. 3.7 and Lemma 3.2(a), implies (a).

(b) For any fixed $x\in S$, a.e. $t\in[0,T]$ (depending on ω), and $0\leq s\leq t$, since

$$\left(e^{\int_s^t\delta c(x_v,\pi_v)dv}\varphi(t,x)\right)'=\delta c(x_t,\pi_t)e^{\int_s^t\delta c(x_v,\pi_v)dv}\varphi(t,x)+e^{\int_s^t\delta c(x_v,\pi_v)dv}\varphi'(t,x),$$

by $\varphi\in\mathbb{B}^1_{V_0,V_1}([0,T]\times S)$ and $T\delta|c(x,a)|\leq M_0V_0(x)$ (using Assumption 3.1(iii)) we have

$$\begin{aligned}\left|\left(e^{\int_s^t\delta c(x_v,\pi_v)dv}\varphi(t,x)\right)'\right|&\leq\frac{M_0}{T}\|\varphi\|_{V_0}V_0(x)e^{\int_s^t|\delta c(x_v,\pi_v)|dv}V_0(x)\\&\quad+\|\varphi'\|_{V_1}e^{\int_s^t|\delta c(x_v,\pi_v)|dv}V_1(x),\\&\leq\left(\frac{\|\varphi\|_{V_0}M_0M_1}{T}+\|\varphi'\|_{V_1}\right)e^{\int_s^t|\delta c(x_v,\pi_v)|dv}V_1(x).\end{aligned}$$

which, together with the same arguments for Eq. 3.3, gives

$$\begin{aligned}&\left|\left(e^{\int_s^t\delta c(x_v(\omega),\pi_v)dv}\varphi(t,x_t(\omega))\right)'\right|\\&\leq\left(\frac{\|\varphi\|_{V_0}M_0M_1}{T}+\|\varphi'\|_{V_1}\right)e^{\int_s^t|\delta c(x_v(\omega),\pi_v)|dv}V_1(x_t(\omega))\end{aligned}\tag{3.8}$$

for a.s. $\omega\in\Omega$ with respect to $\mathbb{P}_x^\pi$, and a.e. $t\in[0,T]$(depending on give ω).

On the other hand, by the Hölder inequality we have

$$\begin{aligned}&\mathbb{E}_\gamma^\pi\left[e^{\int_s^t|\delta c(x_v,\pi_v)|dv}V_1(x_t)|x_s=x\right]\\&\leq\sqrt{\mathbb{E}_\gamma^\pi\left[e^{2\int_s^t|\delta c(x_v,\pi_v)|dv}|x_s=x\right]\mathbb{E}_\gamma^\pi\left[V_1^2(x_t)|x_s=x\right]}\\&\leq\mathbb{E}_\gamma^\pi\left[e^{2\int_s^t|\delta c(x_v,\pi_v)|dv}|x_s=x\right]\mathbb{E}_\gamma^\pi\left[V_1^2(x_t)|x_s=x\right].\end{aligned}\tag{3.9}$$

Furthermore, by the arguments similar to the proof of Eq. 3.1, we also have

$$\mathbb{E}_\gamma^\pi\left[e^{2\int_s^t|\delta c(x_v,\pi_v)|dv}|x_s=x\right]\leq LV_0(x),\tag{3.10}$$

which, together with Lemma 3.1(b) and Eq. 3.9, implies

$$\begin{aligned}&\int_s^T\mathbb{E}_\gamma^\pi\left[e^{\int_s^t|\delta c(x_v,\pi_v)|dv}V_1(x_t)|x_s=x\right]dt\\&\leq LV_0(x)\int_s^T\mathbb{E}_\gamma^\pi\left[V_1^2(x_t)|x_s=x\right]dt\\&\leq TLV_0(x)e^{\rho_1T}\left[V_1^2(x)+\frac{b_1}{\rho_1}\right]<\infty.\end{aligned}\tag{3.11}$$

Also, by Assumptions 3.1 and 3.2, we have

$$\begin{aligned}
&\int_S e^{\int_s^t |\delta c(x_v,\pi_v)|dv} |\varphi(t,y)||q|(dy|x_t,\pi_t)\\
&\le \|\varphi\|_{V_0}[\rho_0 V_0(x_t)+b_0+2M_0' V_0^2(x_t)] e^{\int_s^t |\delta c(x_v,\pi_v)|dv}\\
&\le \|\varphi\|_{V_0}\left[\rho_0 M_1+b_0+2M_0' M_1\right] e^{\int_s^t |\delta c(x_v,\pi_v)|dv} V_1(x_t). \qquad (3.12)
\end{aligned}$$

Thus, by Eqs. 3.8–3.12 we have

$$\begin{aligned}
&\mathbb{E}_\gamma^\pi\left[\int_s^T \left(\left|\left(e^{\int_s^t \delta c(x_v,\pi_v)dv}\varphi(t,x_t)\right)'\right| + \int_S e^{\int_s^t \delta c(x_v,\pi_v)dv}|\varphi(t,y)||q|(dy|x_t,\pi_t)\right) dt|x_s=x\right]\\
&\le L\left[\|\varphi\|_{V_0}M_0M_1+T\|\varphi'\|_{V_1}+T\|\varphi\|_{V_0}(\rho_0M_1+b_0+2M_0'M_1)\right]V_0(x)e^{\rho_1 T}\left[V_1^2(x)+\frac{b_1}{\rho_1}\right]<\infty,
\end{aligned}$$

which, together with Lemma 3.2, verifies (b). □

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

(a) *If there exists* $\varphi \in \mathbb{B}^1_{V_0,V_1}([0,T]\times S)$ *such that*

$$\begin{cases} \varphi'(t,x)+\inf_{a\in A(x)}\left[\delta c(x,a)\varphi(t,x)+\int_S \varphi(t,y)q(dy|x,a)\right]=0, \\ \varphi(T,x)\equiv 1, \end{cases} \qquad (3.13)$$

for each $x\in S$ *and* $t\in E_{(\varphi,x)}$ *with* $m_L(E^c_{(\varphi,x)})=0$, *then*

(a_1) $J(\pi,0,x)\ge\varphi(0,x)$, *for all* $\pi\in\Pi$ *and* $x\in S$, *and*

(a_2) $J(\pi,t,x)\ge\varphi(t,x)$, *for all* $\pi\in\Pi_m^r$ *and* $(t,x)\in[0,T]\times S$.

(b) *For any randomized Markov policy* $\pi\in\Pi_m^r$, $J(\pi,t,x)$ *is a unique solution in* $\mathbb{B}^1_{V_0,V_1}([0,T]\times S)$ *of the following equation*

$$\begin{cases} \varphi'(t,x)+\delta c(x,\pi_t)\varphi(t,x)+\int_S\varphi(t,y)q(dy|x,\pi_t)=0 \\ \varphi(T,x)=1 \end{cases} \qquad (3.14)$$

for each $x\in S$ *and* $t\in E_{(\varphi,x)}$ *with* $m_L(E^c_{(\varphi,x)})=0$.

Proof (a) For each $\omega\in\Omega$, under the conditions for (a) we have

$$\varphi'(t,x)+\int_A \delta c(x,a)\pi(da|\omega,t)\varphi(t,x)+\int_S\int_A \varphi(t,y)q(dy|x,a)\pi(da|\omega,t)\ge 0 \quad (3.15)$$

for all $x\in S$ and $t\in E_{(\varphi,x)}$.

On the other hand, for a.s. $\omega\in\Omega$ (with respect to $\mathbb{P}_x^\pi$), since $\{x_t(\omega),t\in[0,T]\}=:\{x_0,\ldots,x_k\}(x_i\in S, 0\le i\le k)$ for some a finite k (depending on ω) determined by $T_k(\omega)\le T<T_{k+1}(\omega)$ and $m_L(E^c_{(\varphi,x_i)})=0$, by Eq. 3.15 we have

$$\varphi'(t,x_t)+\int_A \delta c(x_t,a)\pi(da|\omega,t)\varphi(t,x_t)+\int_S\int_A \varphi(t,y)q(dy|x_t,a)\pi(da|\omega,t)\ge 0 \quad (3.16)$$

for a.s. $\omega\in\Omega$ and a.e. $t\in[0,T]$ (because of $m_L(\cup_{i=0}^k E^c_{(\varphi,x_i)})=0$).

Let $\bar c(\omega,t,x,\pi):=\int_A \delta c(x_t,a)\pi(da|\omega,t)$. Then, we have, for any $0\le s\le T$,

$$\int_s^T\left[\left(e^{\int_s^t \bar c(\omega,v,x,\pi)dv}\varphi(t,x_t)\right)'+\int_S\int_A q(dy|x_t,a)\pi(da|\omega,t)e^{\int_s^t \bar c(\omega,v,x,\pi)dv}\varphi(t,y)\right]dt\ge 0.$$

The proof of (a_1): For each $\pi \in \Pi$ and $x \in S$, since $\varphi(T, y) = 1$ for all $y \in S$, by Theorem 3.1(a) and Eq. 3.16 we have

$$\begin{aligned}
&\mathbb{E}_x^\pi\left(e^{\int_0^T \int_A \delta c(x_v,a)\pi(da|\omega,v)dv)}\right) - \varphi(0,x)\\
&= \mathbb{E}_x^\pi\left(e^{\int_0^T \int_A \delta c(x_v,a)\pi(da|\omega,v)dv}\varphi(T,x_T)\right) - \varphi(0,x)\\
&= \mathbb{E}_x^\pi\left(e^{\int_0^T \int_A \delta c(x_v,a)\pi(da|\omega,v)dv}\varphi(T,x_T)\right) - \mathbb{E}_x^\pi\left(e^{\int_0^0 \int_A \delta c(x_v,a)\pi(da|\omega,v)dv}\varphi(0,x_0)\right) \geq 0
\end{aligned}$$

and so

$$J(\pi,0,x) = \mathbb{E}_x^\pi\left(e^{\int_0^T \int_A \delta c(x_v,a)\pi(da|\omega,v)dv}\right) \geq \varphi(0,x),$$

which implies (a_1).

Similarly, by Theorem 3.1(b) we see that (a_2) is also true.

(b) If there exists a $\varphi \in \mathbb{B}^1_{V_0,V_1}([0,T]\times S)$ satisfying (3.14), by (a_2) and $\varphi(T,x) \equiv 1$, we have

$$\varphi(s,x) = \mathbb{E}_\gamma^\pi\left[e^{\int_s^T \int_A \delta c(x_v,a)\pi_v(da|x_v)dv}|x_s = x\right] = J(\pi,s,x), \quad s \in [0,T],$$

and so (b) follows. Thus, to complete the proof of (b), it suffices to show the existence of $\varphi \in \mathbb{B}^1_{V_0,V_1}(I \times S)$ satisfying (3.14), while the proof of which is very long and thus will be postponed to Section 4; see Remark 4.2 below.

□

To establish the existence of $\varphi \in \mathbb{B}^1_{V_0,V_1}([0,T]\times S)$ satisfying (3.13), we need Lemma 3.3 below. To state it, we recall some concepts. A subset of a Borel space X is analytic (by Proposition 7.41 in Bertsekas and Shreve 1996) if it is a projection into X of a Borel subset of $X \times Y$ for some uncountable Borel space Y. Then, a function $u(\cdot)$ on X is called upper semianalytic if $\{x \in X : u(x) > r\}$ is an analytic set for each $r \in (-\infty,\infty)$. It is known that each Bore-measurable function is upper semianalytic; see more details in Chapter 7 of Bertsekas and Shreve (1996). Hence, each Borel measurable function such as $c(x,a)$ is upper semianalytic on K.

Lemma 3.3 *Suppose that Assumption 3.1 holds. For any $u(t,x) \in \mathbb{B}_{V_0}([0,T]\times S)$, define a corresponding function $u_*(t,x) : [0,T]\times S \longrightarrow (-\infty,\infty)$ by*

$$u_*(t,x) := \inf_{a\in A(x)}\left\{\delta c(x,a)u(t,x) + \int_S u(t,y)q(dy|x,a)\right\}.$$

Then, the following assertions hold.

(a) *The function $u_*(t,x)$ is upper semianalytic (and hence universally measurable).*

(b) *For every $\varepsilon > 0$, there exists a deterministic Markov policy $f_\varepsilon \in \Pi_m^d$ (depending on ε) such that*

$$\begin{aligned}
&\delta c(x, f_\varepsilon(t,x))u(t,x) + \int_S u(t,y)q(dy|x, f_\varepsilon(t,x))\\
&\leq u_*(t,x) + \varepsilon \;\; \forall\, (t,x) \in [0,\infty)\times S.
\end{aligned}$$

Proof See Lemma 3.3 in Guo et al. (2015b) or (Bertsekas and Shreve 1996, Propositions 7.47 and 7.50). □

4 The existence of optimal Markov policies

In this section, we prove the existence of an optimal Markov policy and a solution to the following optimality equation (4.1) for the finite horizon CTMDPs with the risk-sensitive criterion. The proofs are shown in three steps as follows: 1) consider the case of bounded transition and cost rates, 2) deal with the case of unbounded transition rates but nonnegative costs, and 3) study the case of unbounded transition and cost rates.

Suppose that $f(x)$ is defined on X. Denote $\|f\| := \sup_{x\in X}|f(x)|$. The following results are for the case of the bounded transition and bounded cost rates.

Proposition 4.1 *If $\|q\|$ and $\|c\|$ are finite, then the following assertions hold.*

(a) *There exists a unique φ in $\mathbb{B}^1_{1,1}([0,T]\times S)$ (that is, $V_0(x) = V_1(x) \equiv 1, x \in S$) satisfying the following optimality equation for the risk-sensitive criterion of CTMDPs on the finite horizon:*

$$\begin{cases} \varphi'(t,x) + \inf_{a\in A(x)}[\delta c(x,a)\varphi(t,x) + \int_S \varphi(t,y)q(dy|x,a)] = 0, \\ \varphi(T,x) = 1, \end{cases} \tag{4.1}$$

for each $x \in S$ and $t \in E_{(\varphi,x)}$ with $m_L(E^c_{(\varphi,x)}) = 0$.

(b) *$\varphi(t,x) = J_*(t,x) = \inf_{f\in\Pi^d_m} J(f,t,x)$ for all $(t,x) \in [0,T]\times S$, with $\varphi(t,x)$ as in (a).*

(c) *For any $\varepsilon > 0$, there exists $f_\varepsilon \in \Pi^d_m$ such that $J(f_\varepsilon,t,x) \le \varphi(t,x) + \varepsilon$ for all $(t,x) \in [0,T]\times S$.*

Proof (a) Define the following operator B on $\mathbb{B}_1([0,T]\times S)$ (actually the space of all bounded functions) by

$$B\psi(t,x) := 1 + \int_t^T \inf_{a\in A(x)}\left[\delta c(x,a)\psi(s,x) + \int_S \psi(s,y)q(dy|x,a)\right]ds \tag{4.2}$$

for any $(t,x) \in [0,T]\times S$ and $\psi \in \mathbb{B}_1([0,T]\times S)$.

Then, for each $(t,x) \in [0,T]\times S$, and any $\psi_1, \psi_2 \in \mathbb{B}_1([0,T]\times S)$, from Eq. 4.2 and $q(\{x\}|x,a) + q(S\setminus\{x\})|x,a) \equiv 0$ we obtain

$$\begin{aligned} |B\psi_1(t,x) - B\psi_2(t,x)| &\le \int_t^T \sup_{a\in A(x)}\left[\delta\|c\|\|\psi_1-\psi_2\| + \int_S |q|(dy|x,a)\|\psi_1-\psi_2\|\right]ds \\ &\le (\delta\|c\| + 2\|q\|)\int_t^T \|\psi_1-\psi_2\|ds \\ &= \tilde{L}(T-t)\|\psi_1-\psi_2\| \end{aligned}$$

where $\tilde{L} := \delta\|c\| + 2\|q\| < \infty$. Furthermore, by induction we can prove the following fact:

$$|B^n\psi_1(t,x) - B^n\psi_2(t,x)| \le \tilde{L}^n\frac{(T-t)^n}{n!}\|\psi_1-\psi_2\| \quad \forall\,(t,x) \in [0,T]\times S, n \ge 1. \tag{4.3}$$

Since $\sum_{n=1}^\infty \tilde{L}^n\frac{T^n}{n!}\|\psi_1-\psi_2\| < \infty$, there exists some integer k such that the constant $\beta := \tilde{L}^k\frac{T^k}{k!} < 1$. Thus, by Eq. 4.3 we have $\|B^k\psi_1 - B^k\psi_2\| \le \beta\|\psi_1-\psi_2\|$. Therefore,

B is a k-step contract operator. Thus, there exists a function $\varphi \in \mathbb{B}_1([0,T]\times S)$ such that $B\varphi = \varphi$, that is

$$\varphi(t,x) := 1 + \int_t^T \inf_{a\in A(x)} \left[\delta c(x,a)\varphi(s,x) + \int_S \varphi(s,y)q(dy|x,a)\right] ds \;\forall\, (t,x) \in [0,T]\times S. \quad (4.4)$$

Since $\|q\|$ and $\|c\|$ are finite, by Eq. 4.4 we see that $\varphi \in \mathbb{B}^1_{1,1}([0,T]\times S)$, and thus (a) follows.

(b-c) We are going to prove (b) and (c) together. Since $\|q\| < \infty$ and $\|c\| < \infty$, Assumptions 3.1 and 3.2 are satisfied by taking $V_0 = V_1 \equiv 1$. Thus, it follows from Eq. 4.1 and Theorem 3.2(a) that

$$J(\pi,t,x) \geq \varphi(t,x) \quad \text{for all} \quad \pi \in \Pi^r_m. \qquad (4.5)$$

Moreover, for any $\varepsilon > 0$, since $\varphi \in \mathbb{B}_1([0,T]\times S)$, Lemma 3.3 together with Eq. 4.1 gives the existence of $f_\varepsilon \in \Pi^d_m$, such that, for each $x \in S$ and $t \in E_{(\varphi,x)}$,

$$\begin{cases} \varphi'(t,x) + \delta c(x, f_\varepsilon(t,x))\varphi(t,x) + \int_S \varphi(t,y)q(dy|x, f_\varepsilon(t,x)) \leq \frac{\varepsilon}{T} e^{-T\delta\|c\|}, \\ \varphi(T,x) = 1. \end{cases} \qquad (4.6)$$

Then, as the arguments for Eq. 3.16, by $|c(x,a)| \leq \|c\|$ and Eq. 4.6 we have

$$\left(e^{\int_s^t \delta c(x_v, f_\varepsilon(v,x_v))dv}\varphi(t,x_t)\right)' + \int_S q(dy|x_t, f_\varepsilon(t,x_t))\left(e^{\int_s^t \delta c(x_v, f_\varepsilon(v,x_v)))}\varphi(t,y)\right) \leq \frac{\varepsilon}{T}$$

for every $0 \leq s \leq t$. Thus, by Theorem 3.1(a) we have

$$\begin{aligned} &\mathbb{E}^{f_\varepsilon}_x \left(e^{\int_s^T \delta c(x_v, f_\varepsilon(v,x_v))dv)}\right) - \varphi(s,x) \\ = &\mathbb{E}^{f_\varepsilon}_x \left(e^{\int_s^T \delta c(x_v, f_\varepsilon(v,x_v))dv}\varphi(T,x_T)\right) - \varphi(s,x) \leq \varepsilon \end{aligned}$$

and so

$$J(f_\varepsilon,s,x) \leq \varphi(s,x) + \varepsilon \quad \text{for all } (s,x) \in [0,T]\times S. \qquad (4.7)$$

Therefore, since ε can be arbitrary, by Eqs. 4.5 and 4.7 we have

$$\inf_{\pi\in\Pi^r_m} J(\pi,t,x) = \varphi(t,x) = \inf_{f\in\Pi^d_m} J(f,t,x), \quad \text{and} \quad J(f_\varepsilon,t,x) \leq \varphi(t,x) + \varepsilon,$$

for all $(t,x) \in [0,T]\times S$, and by Eq. 2.6 so (b) and (c) follow.

□

Proposition 4.1 shows the existence of a solution to the optimality equation for the bounded transition and cost rates. To further establish the existence of an optimal Markov policy, we need some conditions below.

Assumption 4.1 (i) For each $x \in S$, $A(x)$ is compact;
(ii) For each $x \in S$ and $D \in \mathcal{B}(S)$, the function $q(D|x,a)$ is continuous in $a \in A(x)$;
(iii) For each $x \in S$, the functions $c(x,a)$ and $\int_S V_0(y)q(dy|x,a)$ are continuous in $a \in A(x)$, with V_0 as in Assumption 3.1.

Remark 4.1 Assumption 4.1 is used for the existence of the minimum points in the optimality equation (4.1).

By Lemma 8.3.7(a) in Hernández-Lerma and Lasserre (1999), under Assumption 4.1 we have the following lemma.

Lemma 4.1 *Under Assumptions 4.1(ii,iii), the function $\int_S q(dy|x,a)u(t,y)$ is continuous in $a \in A(x)$, for every fixed $(t,x) \in [0,T] \times S$ and $u \in \mathbb{B}_{V_0}([0,T] \times S)$.*

Proposition 4.2 *Under Assumption 4.1, if $\|q\| < \infty$ and $\|c\| < \infty$, that is, transition and cost rates are bounded, then the following assertions hold.*

(a) There exists a unique $\varphi(t,x)$ in $\mathbb{B}^1_{V_0,V_1}([0,T] \times S)$ satisfying the optimality equation (4.1).

(b) $\varphi(t,x) = J_*(t,x) = \inf_{f \in \Pi^d_m} J(f,t,x)$ for all $(t,x) \in [0,T] \times S$, with $\varphi(t,x)$ as in (a).

(c) There exists a deterministic Markov policy $f^* \in \Pi^d_m$ such that

$$\varphi'(t,x) + \delta c(x, f^*(t,x))\varphi(t,x) + \int_S \varphi(t,y)q(dy|x, f^*(t,x)) = 0$$

for each $x \in S$ and $t \in E_{(\varphi,x)}$ with $m_L(E^c_{(\varphi,x)}) = 0$.

(d) If, in addition, $c(x,a) \geq 0$ for all $(x,a) \in K$, then $J_*(x,t)$ (and also $\varphi(t,x)$) is decreasing in $t \in [0,T]$ for each fixed $x \in S$.

Proof We only need to prove (c) and (d) since (a) and (b) follow from Proposition 4.1. For the function $\varphi(t,x)$ from (a), when $\varphi'(t,x)$ does not exist for some (t,x), we define

$$\varphi'(t,x) := -\inf_{a \in A(x)} [\delta c(x,a)\varphi(t,x) + \int_S \varphi(t,y)q(dy|x,a)], \tag{4.8}$$

which, together with Eq. 4.1, implies that Eq. 4.8 holds for each $(t,x) \in [0,T] \times S$. Hence, Proposition 7.50 in Bertsekas and Shreve (1996) together with Lemma 4.1 ensures the existence of a Markov policy $f^* \in \Pi^d_m$ such that

$$\begin{cases} \varphi'(t,x)+\delta c(x, f^*(t,x))\varphi(t,x)+\int_S \varphi(t,y)q(dy|x, f^*(t,x)]=0 \quad \forall\, (t,x)\in[0,T] \times S, \\ \varphi(T,x) = 1. \end{cases}$$

Then, as the proofs of (b) and (c) in Proposition 4.1, we see that (c) is true.

(d) Fix any $s,t \in [0,T]$ with $s < t$. Then, for any Markov policy $f \in \Pi^d_m$, we define the corresponding Markov policy f^t_s as follows: for each $x \in S$,

$$f^t_s(v,x) = \begin{cases} f(v+t-s,x) & v \geq s, \\ f(v,x) & \text{otherwise.} \end{cases} \tag{4.9}$$

Then, we have, for each $(v,x) \in [s, s+T-t] \times S$,

$$q(dy|x, f^t_s(v,x)) = q(dy|x, f(v+t-s,x)),\ c(x, f^t_s(v,x)) = c(x, f(v+t-s,x)).$$

Let

$$\begin{aligned} J(f, s \sim t, x) &:= \mathbb{E}^\pi_\gamma \left[e^{\int_s^t \int_A \delta c(x_v, f(v,x_v))dv} | x_s = x \right], \\ J_*(s \sim t, x) &:= \inf_{f \in \Pi^d_m} J(f, s \sim t, x). \end{aligned} \tag{4.10}$$

By the Markov property of $\{x_t, t \geq 0\}$ under any Markov policy f and Eqs. 4.9–4.10, we have $J(f, t \sim T, x) = J(f^t_s, s \sim T+s-t, x)$, and thus $J_*(t \sim T, x) \geq J_*(s \sim T+s-t, x)$; Similarly, we can prove that $J_*(s \sim T+s-t, x) \geq J_*(t \sim T, x)$. Thus, we have $J_*(t \sim T, x) = J_*(s \sim T+s-t, x)$. Moreover, since $c(x,a) \geq 0$ on K, by Eq. 4.10

and $t > s$, we have $J_*(t \sim T, x) = J_*(s \sim T + s - t, x) \leq J_*(s \sim T, x)$, which, together with $J_*(t \sim T, x) = J_*(t, x)$, gives (d). □

Proposition 4.2 shows the existence of an optimal policy under the bounded transition and cost rates. We next extend the results in Proposition 4.2 to the case of unbounded transition rates by approximations from bounded transition and cost rates to unbounded transition rates and nonnegative costs.

Proposition 4.3 *Under Assumptions 3.1, 3.2 and 4.1, if in addition $c(x, a) \geq 0$ for all $(x, a) \in K$, then the following assertions hold.*

(a) *There exists a unique $\varphi(t, x)$ in $\mathbb{B}^1_{V_0,V_1}([0, T] \times S)$ satisfying the optimality equation* (4.1).
(b) *$\varphi(t, x) = J_*(t, x) = \inf_{f \in \Pi^d_m} J(f, t, x)$ for all $(t, x) \in [0, T] \times S$, with $\varphi(t, x)$ as in (a).*
(c) *There exists a Markov policy $f^* \in \Pi^d_m$ such that*

$$\varphi'(t, x) + \delta c(x, f^*(t, x))\varphi(t, x) + \int_S \varphi(t, y) q(dy|x, f^*(t, x)) = 0$$

for each $x \in S$ and $t \in E_{(\varphi,x)}$ with $m_L(E^c_{(\varphi,x)}) = 0$.

Proof We only proof (a). This is because (b) and (c) can be proved as the same arguments of (b)-(c) of Proposition 4.2.

The main thread of our proof is as follows.

First, we construct a series of models $\mathcal{M}^+_n$ which the transition rates and costs are all bounded. By Proposition 4.2, there exists a decreasing (with respect to t) function $\varphi_n(t, x) \in \mathbb{B}^1_{1,1}([0, T] \times S)$ which is the value function of $\mathcal{M}^+_n$.

Second, we prove that the limit of $\varphi_n(t, x)$ exists, denoted by $\varphi(t, x)$.

Third, we prove that $\varphi(t, x)$ is the value function of the original model.

Now, we construct $\mathcal{M}^+_n$ first. under Assumption 3.1 (iii), we have

$$1 \leq e^{2T\delta c(x,a)} \leq M_0 V_0(x), \text{ (i.e., } 0 \leq c(x, a) \leq \frac{1}{T\delta} \ln \sqrt{M_0 V_0(x)}) \text{ for all } (x, a) \in K.$$

For each $n \geq 1$, let $A_n(x) := A(x)$ for $x \in S$, $K_n := \{(x, a)|x \in S, a \in A_n(x)\}$, and $S_n := \{x \in S|V_0(x) \leq n\}$. Moreover, for each $x \in S$, $a \in A_n(x)$, let

$$q_n(dy|x, a) := \begin{cases} q(dy|x, a) & \text{if } x \in S_n, \\ 0 & \text{if } x \notin S_n; \end{cases} \tag{4.11}$$

$$c^+_n(x, a) := \begin{cases} c(x, a) \wedge \min\{n, \frac{1}{T\delta} \ln \sqrt{M_0 V_0(x)}\} & \text{if } x \in S_n, \\ 0 & \text{if } x \notin S_n. \end{cases} \tag{4.12}$$

Fix any $n \geq 1$. By Eq. 4.11, it is obvious that the $q_n(dy|t, x, a)$ denotes indeed transition rates on S, which are *conservative* and *stable*. By Eq. 4.12, recall $T > 0, \delta > 0, M_0 \geq 1, V_0(x) \geq 1$ and $c(x, a) \geq 0$, $c^+_n(x, a) \geq 0$ for $n \geq 1$. Then, we obtain a sequence of models $\{\mathcal{M}^+_n\}$:

$$\mathcal{M}^+_n := \left\{S, A, (A_n(x), x \in S), c^+_n(x, a), q_n(\cdot|x, a)\right\},$$

for which the transition rates $q_n(dy|x, a)$ and costs $c^+_n(x, a)$ are all bounded (by Assumption 3.1 and Eqs. 4.11–4.12). In the following arguments, any quality with respect to $\mathcal{M}^+_n$ is labeled by a lower n, such as the risk-sensitive criterion $J_n(f, t, x)$ of a Markov policy f and the value function $J_n(t, x) := \inf_{f \in \Pi^d_m} J_n(f, t, x)$.

Obviously, Assumptions 3.1, 3.2 and 4.1 still hold for each model $\mathcal{M}_n^+$. Thus, for each $n \geq 1$, it follows from Proposition 4.2 that there exists $\varphi_n(t,x) \in \mathbb{B}^1_{1,1}([0,T]\times S)$ satisfying (4.1) for the corresponding $\mathcal{M}_n^+$, that is,

$$\begin{cases} \varphi_n'(t,x) + \inf_{a\in A_n(x)}[\delta c_n^+(x,a)\varphi_n(t,x) + \int_S \varphi_n(t,y)q_n(dy|x,a)] = 0, \\ \varphi_n(T,x) = 1. \end{cases} \tag{4.13}$$

for all $x \in S$ and $t \in E_{(\varphi_n,x)}$ with $m_L(E^c_{(\varphi_n,x)}) = 0$. Furthermore, since $c_n^+(x,a) \geq 0$, by Proposition 4.2 (d), $\varphi_n(t,x))$ is decreasing in $t \in [0,T]$ for each fixed $x \in S$.

Then, Proposition 7.50 in Bertsekas and Shreve (1996) together with Lemma 4.1 and Eq. 4.13 gives the existence of a Markov policy $f_n \in \Pi_m^d$ such that,

$$\begin{cases} \varphi_n'(t,x) + \delta c_n^+(x, f_n(t,x))\varphi_n(t,x) + \int_S \varphi_n(t,y)q_n(dy|x, f_n(t,x) = 0 \quad \forall\, x \in S, \\ \varphi_n(T,x) = 1 \end{cases} \tag{4.14}$$

for all $x \in S$ and $t \in E_{(\varphi_n,x)}$ with $m_L(E^c_{(\varphi_n,x)}) = 0$.

Also, by Eq. 4.12 we have $e^{2T\delta c_n^+(x,a)} \leq M_0 V_0(x)$ for all $x \in S$ and $n \geq 1$. Then, using Lemma 3.1 and Theorem 3.2(a_2) with $V_0 = V_1 \equiv 1$ and Lemma 3.1, from Eq. 4.14 we have

$$e^{-LV_0(x)} \leq \varphi_n(t,x) = J_n(f_n,t,x) \leq LV_0(x) \quad \forall\, n \geq 1. \tag{4.15}$$

Moreover, since $\varphi_n(t,x) \geq 0$ and $c_n^+(x,a) \geq c_{n-1}^+(x,a)$ for all $(x,a) \in K$, by Eqs. 4.11 and 4.14 as well as Proposition 4.2(d), we have, for all $x \in S$, $t \in E_{(\varphi_n,x)}$ and $n \geq 2$

$$\begin{cases} \varphi_n'(t,x) + \delta c_{n-1}^+(x, f_n(t,x))\varphi_n(t,x) + \int_S \varphi_n(t,y)q_{n-1}(dy|x, f_n(t,x)) \leq 0, \ x \in S_{n-1}, \\ \varphi_n'(t,x) + \delta c_{n-1}^+(x, f_n(t,x))\varphi_n(t,x) + \int_S \varphi_n(t,y)q_{n-1}(dy|x, f_n(t,x)) = \varphi_n'(t,x) \leq 0, \ x \notin S_{n-1} \\ \varphi_n(T,x) = 1, \end{cases}$$

which, together with Theorem 3.2(a_2) with $V_0 = V_1 \equiv 1$, implies that $J_{n-1}(f_n,t,x) \leq \varphi_n(t,x)$ for all $(x,t) \in [0,T] \times S$. Therefore, we have $\varphi_{n-1}(t,x) \leq J_{n-1}(f_n,t,x) \leq \varphi_n(t,x)$, that is, the sequence $\{\varphi_n, n \geq 1\}$ is nondecreasing in $n \geq 1$, and thus the limit

$$\varphi(t,x) := \lim_{n\to\infty} \varphi_n(t,x) \tag{4.16}$$

exists for each $(t,x) \in [0,T] \times S$.

Since $\varphi_n(t,x)$ is decreasing in $t \in [0,T]$, $\varphi(t,x)$ is decreasing in $t \in [0,T]$, for each fixed $x \in S$. Therefore, $\varphi(t,x)$ is differential in a.e. $t \in [0,T]$ (for each fixed $x \in S$).

Let, for every $n \geq 1$ and $(t,x) \in [0,T] \times S$,

$$H_n(t,x) := \inf_{a\in A(x)} \left[\delta c_n^+(x,a)\varphi_n(t,x) + \int_S \varphi_n(t,y)q_n(dy|x,a)\right],$$
$$H(t,x) := \inf_{a\in A(x)} \left[\delta c(x,a)\varphi(t,x) + \int_S \varphi(t,y)q(dy|x,a)\right].$$

We next show that $\lim_{n\to\infty} H_n(t,x) = H(t,x)$ for each $(t,x) \in [0,T] \times S$.

Indeed, for any fixed $(t,x) \in [0,T] \times S$, there exists $n_0 \geq 1$ such that $(t,x) \in [0,T] \times S_{n_0}$, and then $q_n(dy|x,a) = q(dy|x,a)$ for all $n \geq n_0$ and $\lim_{n\to\infty} c_n^+(x,a) = c(x,a)$ for all $a \in A(x)$. Thus, by Lemma 8.3.7 in Hernández-Lerma and Lasserre (1999) and Eq. 4.15 we have

$$\begin{aligned}
&\limsup_{n\to\infty} H_n(t,x)\\
&\le \limsup_{n\to\infty}\left[\delta c_n^+(x,a)\varphi_n(t,x)+\int_S \varphi_n(t,y)q(dy|x,a)\right]\\
&\le \delta c(x,a)\varphi(t,x)+\int_S \varphi(t,y)q(dy|x,a), \quad \text{for all } a \in A(x).
\end{aligned}$$

Hence,

$$\limsup_{n\to\infty} H_n(t,x) \le \inf_{a\in A(x)}\left[\delta c(x,a)\varphi(t,x)+\int_S \varphi(t,y)q(dy|x,a)\right]. \tag{4.17}$$

On the other hand, note that $\liminf_{n\to\infty} H_n(t,x) = \lim_{m\to\infty} H_{n_m}(t,x)$ for some subsequence $\{n_m, m \ge 1\}$ of $\{n, n \ge 1\}$. For each $m \ge 1$, under Assumption 4.1, the measurable selection theorem (e.g. Proposition 7.50 in Bertsekas and Shreve 1996) together with Lemma 4.1 ensures the existence of $f_{n_m} \in \Pi_m^d$ such that

$$\begin{aligned}
H_{n_m}(t,x) &= \inf_{a\in A(x)}[\delta c_{n_m}^+(x,a)\varphi_{n_m}(t,x)+\int_S \varphi_{n_m}(t,y)q(dy|x,a)]\\
&= \delta c_{n_m}^+(x,f_{n_m}(t,x))\varphi_{n_m}(t,x)+\int_S \varphi_{n_m}(t,y)q(dy|x,f_{n_m}(t,x)).
\end{aligned} \tag{4.18}$$

Since $f_{n_m}(t,x) \in A(x)$ for all $m \ge 1$ and $A(x)$ is compact, there exists a subsequence $\{f_{n_{m_k}}(t,x), k \ge 1\}$ of $\{f_{n_m}(t,x), m \ge 1\}$ and $a(t,x) \in A(x)$ (depending on (t,x)) such that $f_{n_{m_k}}(t,x) \to a(t,x)$ as $k \to \infty$. Thus, since $\lim_{m\to\infty} H_{n_m}(t,x) = \lim_{k\to\infty} H_{n_{m_k}}(t,x)$, using Assumption 4.1, by Lemma 8.3.7 in Hernández-Lerma and Lasserre (1999) and Eq. 4.18 we have

$$\begin{aligned}
\liminf_{n\to\infty} H_n(t,x) &= \lim_{k\to\infty} H_{n_{m_k}}(t,x)\\
&= \lim_{k\to\infty}\left[\delta c_{n_{m_k}}^+(x,f_{n_{m_k}}(t,x))\varphi_{n_{m_k}}(t,x)+\int_S \varphi_{n_{m_k}}(t,y)q(dy|x,f_{n_{m_k}}(t,x))\right]\\
&= \delta c(x,a(t,x))\varphi(t,x)+\int_S \varphi(t,y)q(dy|t,x,a(t,x))\\
&\ge \inf_{a\in A(x)}\left[\delta c(x,a)\varphi(t,x)+\int_S \varphi(t,y)q(dy|x,a)\right],
\end{aligned}$$

which, together with Eq. 4.17, implies that $\lim_{n\to\infty} H_n(t,x) = H(t,x)$. Thus, by Eq. 4.13 we have

$$\varphi(t,x) = 1+\int_t^T \inf_{a\in A(x)}\left[\delta c(x,a)\varphi(s,x)+\int_S \varphi(s,y)q(dy|x,a)\right]ds. \tag{4.19}$$

Since $\varphi(t,x)$ is the integral of a measurable function, it is a absolutely continuous function. Therefore, we prove that $\varphi(t,x)$ is differential in a.e. $t \in [0,T]$ (for each fixed $x \in S$) again. We can verify that $\varphi(t,x)$ satisfies (4.1). To show $\varphi(t,x) \in \mathbb{B}^1_{V_0,V_1}([0,T]\times S)$, since

$\varphi(t,x) \in \mathbb{B}_{V_0}([0,T]\times S)$ (by Eqs. 4.15–4.16), the rest verifies that $\varphi'(t,x)$ is V_1-bounded. Indeed, since $T|\delta c(x,a)| \le e^{T\delta|c(x,a)|} \le M_0V_0(x)$, from Eq. 4.19 we have

$$\begin{aligned}|\varphi'(t,x)| &\le \frac{M_0}{T}\|\varphi\|_{V_0}V_0(x)V_0(x) + \|\varphi\|_{V_0}[\rho_0V_0(x)+b_0+2V_0(x)q^*(x)]\\ &\le \|\varphi\|_{V_0}\left[\frac{M_0}{T}V_0^2(x)+\rho_0V_0(x)+b_0+2M_0'V_0^2(x)\right]\\ &\le \|\varphi\|_{V_0}\left[\frac{M_1M_0}{T}+\rho_0M_1+b_0+2M_0'M_1\right]V_1(x)\end{aligned}$$

which implies that $\varphi(t,x)$ is in $\mathbb{B}^1_{V_0,V_1}([0,T]\times S)$, and thus (a) is proved. □

Next, we use Proposition 4.3 to prove our main results by approximation from nonnegative cost rates to the cost rates that may be unbounded from above and from below.

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

(a) There exists a unique $\varphi(t,x)$ in $\mathbb{B}^1_{V_0,V_1}([0,T]\times S)$ satisfying the optimality equation (4.1).

(b) $\varphi(t,x) = \inf_{\pi\in\Pi^r_m} J(\pi,t,x) = \inf_{f\in\Pi^d_m} J(f,t,x)$ for all $(t,x)\in[0,T]\times S$, with $\varphi(t,x)$ as in (a).

(c) There exists a Markov policy $f^*\in\Pi^d_m$ such that

$$\varphi'(t,x)+\delta c(x,f^*(t,x))\varphi(t,x)+\int_S\varphi(t,y)q(dy|x,f^*(t,x))=0$$

for each $x\in S$ and $t\in E_{(\varphi,x)}$ with $m_L(E^c_{(\varphi,x)})=0$, and f^* is optimal.

Proof We only prove (a) and the optimality of the policy f^* since the others can be proved as (b) and (c) of Proposition 4.3. For each $n\ge 1$, define c_n on K as follows: for each $(x,a)\in K$,

$$c_n(x,a) := \max\{-n, c(x,a)\},$$

which implies that $\lim_{n\to\infty}c_n(x,a)=c(x,a)$ and $k_n(x,a):=c_n(x,a)+n\ge 0$ for each $(x,a)\in K$ and $n\ge 1$. Moreover, it follows from Assumption 3.1(iii) that

$$-\frac{1}{T\delta}\ln\sqrt{M_0V_0(x)} \le \max\{-n,-\frac{1}{T\delta}\ln\sqrt{M_0V_0(x)}\} \le c_n(x,a) \le \frac{1}{T\delta}\ln\sqrt{M_0V_0(x)}. \tag{4.20}$$

Thus, $e^{2T\delta k_n(x,a)} \le e^{2Tn\delta}M_0V_0(x)$ for all $(x,a)\in K$ for all $n\ge 1$, and so Assumptions 3.1 (with M_0 replaced by $e^{2Tn\delta}M_0$), 3.2 and 4.1 still hold for each model $\mathcal{N}_n$ defined by

$$\mathcal{N}_n := \{S,(A(x),x\in S),k_n(x,a),q(\cdot|x,a)\}.$$

For any real-valued Borel measurable function u on K, let

$$J_u(x,t) := \inf_{f\in\Pi^d_m}\mathbb{E}^\pi_\gamma\left[e^{\delta\int_t^T\int_A u(x_t,f(t,x_t))dt}|x_t=x\right]$$

provided the integral exists. Then, for each $n\ge 1$, since $k_n(x,a)\ge 0$, by Proposition 4.3(b) we have J_{k_n} is in $\mathbb{B}^1_{V_0,V_1}([0,T]\times S)$ and satisfies

$$\begin{cases} J'_{k_n}(t,x)+\inf_{a\in A(x)}[\delta k_n(x,a)J_{k_n}(t,x)+\int_S J_{k_n}(t,y)q(dy|x,a)]=0\\ J_{k_n}(T,x)=1.\end{cases} \tag{4.21}$$

for all $x \in S$ and $t \in E_{(J_{k_n},x)}$.

Moreover, since $J_{k_n}(t,x) = J_{c_n+n}(t,x) = J_{c_n}(t,x)e^{\delta(T-t)n}$, by Eq. 4.21 we derive that

$$\begin{cases} J'_{c_n}(t,x) + \inf_{a\in A(x)}[\delta c_n(x,a)J_{c_n}(t,x) + \int_S J_{c_n}(t,y)q(dy|x,a)] = 0 \\ J_{c_n}(T,x) = 1. \end{cases}$$

This is

$$J_{c_n}(t,x) = 1 + \int_t^T \inf_{a\in A(x)}[\delta c_n(x,a)J_{c_n}(s,x) + \int_S J_{c_n}(s,y)q(dy|x,a)]ds. \quad (4.22)$$

On the other hand, for each $(t,x) \in [0,T] \times S$, it follows from Eq. 4.20 and Lemma 3.1(d) that

$$|J_{c_n}(t,x)| \le LV_0(x), \quad n \ge 1. \quad (4.23)$$

Since $c_n(x,a)$ is decreasing in $n \ge 1$, and so is the corresponding value functions $J_{c_n}(t,x)$. Therefore, the limit $\varphi(t,x) := \lim_{n\to\infty} J_{c_n}(t,x)$ exists for each $(t,x) \in [0,T]\times S$. Then, as the arguments for Proposition 4.3 with $\varphi_n(t,x)$ replaced with $J_{c_n}(t,x)$ here, from Eqs. 4.22 and 4.23 we can see that (a) is also true.

Moreover, by Eq. 4.1 and (c), using Theorem 3.2 we see that f^* is optimal. □

Remark 4.2 For the given $\pi_t(da|x) \in \Pi_m^r$, to show the existence of a $\varphi \in \mathbb{B}^1_{V_0,V_1}([0,T]\times S)$ satisfying (3.14), we modify the operator B in Eq. 4.2 as the following B^π:

$$B^\pi\psi(t,x) := 1 + \int_t^T \left[\delta c(x,\pi_s)\psi(s,x) + \int_S \psi(s,y)q(dy|x,\pi_s)\right]ds$$

for all $(t,x) \in [0,T] \times S$. Then, a similar argument as in the proof of Proposition 4.3(a) gives the existence of a $\varphi \in \mathbb{B}^1_{V_0,V_1}([0,T] \times S)$ satisfying (3.14).

5 An example

Recall that Assumptions 3.1 and 4.1 above are the generalization of the corresponding ones in Guo and Hernández-Lerma (2009), Guo et al. (2012), Guo and Piunovskiy (2011), Piunovskiy and Zhang (2011), and Prieto-Rumeau and Hernández-Lerma (2012). Hence, they are satisfied for all the examples and hypotheses in these references. To further illustrate the main results here, we next consider the risk-sensitive optimality problem of case flow in Guo et al. (2015a) for CTMDPs on the mean-variance criteria.

Example 5.1 (Risk-sensitive controlled problems of cash flow) Consider a continuous-time controlled problem of cash flow in an economic market with an amount of the cash as a state, and thus the corresponding state space is $S := (-\infty, +\infty)$. When the current state of cash flow is at $x \in S$, a decision-maker withdraws money with the amount $-a$ (if $a < 0$) or takes a supply of money with the amount a for $a \ge 0$, where a is regarded an action. When the current state is at $x \in S$ and an action $a \in A(x)$ is chosen, the two things happen: 1) a cost is incurred at rate $c(x,a)$; and 2) the amount x of cash is assumed to keep invariable for an exponential-distributed random time with parameter $\lambda(x,a) \ge 0$, and then jump to other states with the normal distribution $N(x,\sigma^2)$ for some constant $\sigma > 0$. Therefore, the transition rates of cash flow is represented by

$$q(D|x,a) := \lambda(x,a)\left[\frac{1}{\sqrt{2\pi}\sigma}\int_D e^{-\frac{(y-x)^2}{2\sigma^2}}dy - \delta_x(D)\right] \quad \text{for each } D \in \mathcal{B}(S). \quad (5.1)$$

For this cash flow model, the decision maker wishes to minimize the risk-sensitive costs on a given T horizon over all policies.

To ensure the existence of an optimal policy for the cash flow model, we consider the following hypotheses:

(A_1) $\lambda(x,a) \leq M(x^2+1)$ and $|c(x,a)| \leq \frac{1}{2\delta T}[M+\ln(x^2+1)]$ for all $x \in S, a \in A(x)$ with some positive constant M, where the constants δ and T are as before;
(A_2) $A(x)$ is assumed to be a compact set of a Borel space A for each $x \in S$;
(A_3) $\lambda(x,a)$ and $c(x,a)$ are Borel measurable on K and continuous in $a \in A(x)$ for each fixed $x \in S$.

Under the above conditions, we have the following fact.

Proposition 5.1 *Under the hypotheses A_1–A_3, Example 5.1 satisfies the Assumptions 3.1, 3.2 and 4.1, and hence (by Theorem 4.1) there exists an optimal Markov policy.*

To verify the conditions required in Theorem 4.1, let

$$V_0(x) := 1 + x^2,\, V_1(x) := 1 + x^4 \quad \forall\, x \in S.$$

Since $\frac{1}{\sqrt{2\pi}\sigma}\int_S (y-x)^{2k+1} e^{-\frac{(y-x)^2}{2\sigma^2}} dy = 0$ and $\frac{1}{\sqrt{2\pi}\sigma}\int_S (y-x)^{2k} e^{-\frac{(y-x)^2}{2\sigma^2}} dy = 1\cdot 3\cdots(2k-1)\sigma^{2k}$ for all $k = 0, 1, \ldots$, using (5.1) and hypothesis A_1, a directive calculation gives

$$\begin{aligned}\int_S V_0(y)q(dy|x,a) &= \lambda(x,a)\left[\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(y^2+1)e^{-\frac{(y-x)^2}{2\sigma^2}}dy - (x^2+1)\right]\\ &= \lambda(x,a)\sigma^2 \leq M\sigma^2 V_0(x);\end{aligned}$$

$$\begin{aligned}\int_S V_1^2(y)q(dy|x,a) &= \lambda(x,a)\left[\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(y^4+1)^2 e^{-\frac{(y-x)^2}{2\sigma^2}}dy - (x^4+1)^2\right]\\ &= \lambda(x,a)\Big(105\sigma^8+420x^2\sigma^6+210x^4\sigma^4+6\sigma^4+12\sigma^2x^2+28x^6\sigma^2\Big)\\ &\leq 420\lambda(x,a)\left(\sigma^8+\sigma^6+\sigma^4+\sigma^2\right)\left(x^6+x^4+x^2+1\right);\\ &\leq 420\lambda(x,a)\left(\sigma^8+\sigma^6+\sigma^4+\sigma^2\right)\left(3x^6+3\right);\\ &\leq 1260M\left(\sigma^8+\sigma^6+\sigma^4+\sigma^2\right)\left(x^6+1\right)(1+x^2);\\ &\leq 3780M\left(\sigma^8+\sigma^6+\sigma^4+\sigma^2\right)\left(x^4+1\right)^2;\\ &= 3780M\left(\sigma^8+\sigma^6+\sigma^4+\sigma^2\right)V_1^2(x).\end{aligned}$$

Thus, the hypotheses A_1–A_3 imply the Assumptions 3.1, 3.2 and 4.1, and then Theorem 4.1 gives the existence of an optimal Markov policy.

Remark 5.1 In this example, the cost $c(x,a)$ are allowed to be unbounded from above and below, and the transition rates $q(dy|x,a)$ can be unbounded. Thus, some of the conditions in Ghosh and Saha (2014), Jaskiewicz (2007), Kumar and Chandan (2015), Kumar and Chandan (2013), and Wei (2016) for CTMDPs on the risk-sensitive criteria fails to hold for this example because the transition and cost rates are all bounded in Ghosh and Saha (2014), Jaskiewicz (2007), Kumar and Chandan (2013, 2015) and Wei (2016).

Acknowledgments Research supported by the National Natural Science Foundation of China (Grant No. 60171009, Grant No. 61673019, Grant No. 11931018).

References

Anantharam V, Borkar VS (2017) A variational formula for risk-sensitive reward. SIAM J Control Optim 55:961–988

Basu A, Ghosh MK (2014) Zero-sum risk-sensitive stochastic games on a countable state space. Stochastic Process Appl 124:961–983

Baüerle N, Rieder U (2014) More risk-sensitive Markov decision processes. Math Oper Res 39:105–120

Baüerle N, Rieder U (2017) Zero-sum risk-sensitive stochastic games. Stochastic Process Appl 127:622–642

Bertsekas D, Shreve S (1996) Stochastic optimal control: the discrete-time case. Academic Press, Inc

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

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

Feinberg EA, Mandava M, Shiryaev AN (2014) On solutions of Kolmogorov's equations for nonhomogeneous jump Markov processes. J Math Anal Appl 411:261–270

Guo X (2007) Continuous–time Markov decision processes with discounted rewards: the case of Polish spaces. Math Oper Res 32:73–87

Guo X, Piunovskiy A (2011) Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates. Math Oper Res 36:105–132

Guo X, Song X (2011) Discounted continuous-time constrained Markov decision processes in Polish spaces. Ann Appl Probab 21:2016–2049

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

Guo X, Huang Y, Song X (2012) Linear programming and constrained average optimality for general continuous-time Markov decision processes in history-dependent policies. SIAM J Control Optim 50:23–47

Guo X, Huang XX, Zhang Y (2015a) On the first passage g-mean-variance optimality for discounted continuous-time Markov decision processes. SIAM J Control Optim 53:1406–1424

Guo X, Huang XX, Huang Y (2015b) 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

Huang Y (2018) Finite horizon continuous-time Markov decision processes with mean and variance criteria. Discret Event Dyn Syst 28(4):539–564

Huo H, Zou X, Guo X (2017) The risk probability criterion for discounted continuous-time Markov decision processes. Discret Event Dyn Syst 27(4):675–699

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

Kitaev MY, Rykov V (1995) Controlled queueing systems. CRC Press, New York

Kumar KS, Chandan P (2013) Risk-sensitive control of jump process on denumerable state space with near monotone cost. Appl Math Optim 68:311–331

Kumar KS, Chandan P (2015) Risk-sensitive control of continuous-time Markov processes with denumerable state space. Stoch Anal Appl 33:863–881

Piunovskiy A, Zhang Y (2011) Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. SIAM J Control Optim 49:2032–2061

Prieto-Rumeau T, Hernández-Lerma O (2012) Selected topics in continuous-time controlled Markov chains and Markov games. Imperial College Press, London

Wei QD (2016) Continuous-time Markov decision processes with risk-sensitive finite-horizon cost criterion. Math Meth Oper Res 84:461–487

Wei Q, Chen X (2017) Average cost criterion induced by the regular utility function for continuous-time Markov decision processes. Discret Event Dyn Syst 27(3):501–524

Xia L (2014) Event-based optimization of admission control in open queueing networks. Discret Event Dyn Syst 24(2):133–151

Xia L (2018) Variance minimization of parameterized Markov decision processes. Discret Event Dyn Syst 28:63–81

Yushkevich AA (1977) Controlled Markov models with countable state and continuous time. Theory Probab Appl 22:215–235

Zhang Y (2017) Continuous-time Markov decision processes with exponential utility. SIAM J Control Optim 55:2636–2660

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Dr. Xianping Guo received the Ph.D. degree in probability and statistics from the Central South University, Changsha, China, in 1996. He has published many papers in international journals such as the Annals of Applied Probability, SIAM Journal on Optimization, SIAM Journal on Control and Optimization, IEEE Transactions on Automatic, and Automatica, and also published a book titled Continuous-Time Markov Decision Processes in Springer (co-authored with Professor Onesimo Hernandez-Lerma). He has been an Editor of Advances in Applied Probability, Journal of Applied Probability, and of Science China Mathematics. He is currently with Sun Yat-Sen University, Guangzhou, China, where he was appointed as a Professor in 2002. His research interests include stochastic optimality, and stochastic games.

Junyu Zhang received the B.S. degree in statistics and probability from Peking University, China, in 1999, the M.S. degree from Academy of Mathematics and Systems Science, Chinese Academic of Science, in 2002, and the Ph.D. degree in electrical and electronic engineering from the Hong Kong University of Science and Technology in 2006. Since 2006, she has been an associate Professor in Mathematics at the Sun Yat-sen University, Guangzhou, China. Her research interests include Markov decision processes, stochastic optimization, and discrete event dynamic systems.