Finite horizon risk-sensitive continuous-time Markov decision processes with unbounded transition and cost rates

Xin Guo[*], Qiuli Liu[†] and Yi Zhang[‡]

Abstract: We consider a risk-sensitive continuous-time Markov decision process over a finite time duration. Under the conditions that can be satisfied by unbounded transition and cost rates, we show the existence of an optimal policy, and the existence and uniqueness of the solution to the optimality equation out of a class of possibly unbounded functions, to which the Feynman-Kac formula was also justified to hold.

Keywords: Continuous-time Markov decision processes. Risk-sensitive criterion. Optimality Equation.

AMS 2000 subject classification: Primary 90C40, Secondary 60J75

1 Introduction

This paper considers a continuous-time Markov decision process (CTMDP) in a denumerable state space, where the performance is measured by the expectation of the exponential utility of the total cost over a fixed finite time horizon with possibly unbounded transition and cost rates. Such problems are often called risk-sensitive because they take both the first order (expectation) and higher order moments. We give the following formal justifications as in [3, 5]. Let X be the (possibly random) cost, and so the concerned performance measure is $E[e^{\theta X}]$, where $\theta > 0$ is a fixed constant. Let $ceq(X) := \frac{1}{\theta}\ln(E[e^{\theta X}])$ be a (deterministic) constant such that $E[e^{\theta X}] = e^{\theta ceq(X)}$. (For simplicity, assume all the involved expectations are finite.) Then applying the Taylor expansion around $E[X]$:

$$e^{\theta ceq(X)} \approx e^{\theta E[X]} + \theta e^{\theta E[X]}(ceq(X) - E[X]);$$
$$e^{\theta X} \approx e^{\theta E[X]} + \theta e^{\theta E[X]}(X - E[X]) + \frac{1}{2}\theta^2 e^{\theta E[X]}(X - E[X])^2$$

so that $e^{\theta ceq(X)} = E[e^{\theta X}] \approx e^{\theta E[X]} + \frac{1}{2}\theta^2 e^{\theta E[X]} Var(X)$ by taking expectation on the both sides of the second equality in the above. Comparing this with the first equality in the above, we see $ceq(X) - E[X] \approx \frac{1}{2}\theta Var(X) \geq 0$. Thus, the performance of $E[e^{\theta X}]$ takes into account both $E[X]$ and $Var(X)$ compared to the case of linear utility where only $E[X]$ is counted. We also mention that $ceq(X)$ is the so called certainty equivalent of X, which can be interpreted as the deterministic price the decision maker is willing to pay in exchange for the random cost, and so $ceq(X) - E[X] \geq 0$ means the decision maker is risk-averse. Below, we will regard $\theta = 1$ without loss of generality. The

[*] Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, U.K.. E-mail: X.Guo21@liv.ac.uk.

[†] School of Mathematical Sciences, South China Normal University, Guangzhou, 510631. China. E-mail: liuql2007@aliyun.com.

[‡] Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, U.K.. E-mail: yi.zhang@liv.ac.uk.

CTMDP with a linear utility is called risk-neutral. Risk-sensitive and risk-neutral problems might admit quite different optimality results in general. For example, in a model with finite state and action space, there is always an optimal deterministic stationary policy for discounted risk-neutral CTMDPs, whereas this is not the case for the risk sensitive counterpart, see [5].

Risk-sensitive Markov decision processes (in discrete-time) have been studied intensively since 1970s, with one of the pioneering works being [11], and a recent and updated work being [1], to which the interested reader is referred for more references. Compared to the discrete-time framework, there have been fewer works on risk-sensitive CTMDPs, also known as controlled Markov pure jump processes. An early work on this topic seems to be [18], which obtained verification theorems and solved in closed-form meaningful examples of problems over a fixed time duration. In the recent years, there have been reviving interests in risk-sensitive CTMDPs, see [5, 6, 21] for problems with a finite horizon, [23] for problems over an infinite horizon, [5, 16, 22] for problems with average criteria, and [2] for an optimal stopping problem with a more general utility function than the exponential one.

In greater detail, the CTMDP considered in [5] is with bounded transition and cost rates. In [21], the boundedness on the transition rate was relaxed and replaced by a drift-type condition, but the cost rate was still assumed to be bounded. Both papers followed the same line of reasoning: they showed the existence of a solution to the optimality equation, and then showed that the solution coincides with the value function of the problem by applying the Feynman-Kac formula. In Section 7 of [21], the author mentioned that following his method it was unclear how to relax the boundedness assumption on the cost rate at that time, as a suitable version of the Feynman-Kac formula must be established first. The present paper provides a response to this. In greater detail, the main contributions are the following. We provide conditions that allow unbounded transition and (not necessarily nonnegative) cost rates, under which a suitable version of the Feynman-Kac formula was established, and we show that the value function is the unique solution out of a large enough class of functions (possibly unbounded with unbounded derivatives with respect to time) to the optimality equation. It is important for practical applications to consider models with unbounded transition and cost rates. We illustrate this with an example of controlled $M/M/\infty$ queueing system. Compared with [5, 21], which concentrated on Markov policies, we consider a more general class of policies. When the cost rate is nonnegative, a different method was followed in [6], which is not based on the Feynman-Kac formula. If the cost rate is nonnegative, then the conditions on the transition and cost rates in [6] are weaker than in the present paper. Moreover, in that general setup of [6], the value function is generally not the unique solution to the optimality equation. In this sense, the present paper also complements [6].

The rest of the paper is organized as follows. In Section 2 we describe the optimal control problem under consideration. Section 3 contains preliminary results, where we establish a version of the Feynman-Kac formula. The optimality results are proved in Section 4. This paper is finished with a conclusion in Section 5.

2 Model description

Notation: For a Borel space X endowed with the Borel σ-algebra $\mathcal{B}(X)$, we denote by $\mathbb{C}_b(X)$ the space of all bounded continuous functions on X with the norm $\|u\| := \sup_{x\in X} |u(x)|$. Throughout this paper, measurability is understood in the Borel sense.

We consider the CTMDP model $\mathcal{M} := \{S, A, A(\cdot,\cdot), q, c, g\}$ consisting of the following elements. The state space S is a denumerable set, endowed with the discrete topology. The action space A is a (nonempty) Borel space. The multifunction $(t, i) \in [0, \infty) \times S \to A(t, i) \in \mathcal{B}(A)$ specifies the set of admissible action spaces given the current time and state, and is assumed to be with a measurable graph $K := \{(t, i, a) \in [0, \infty) \times S \times A : a \in A(t, i)\}$, containing the graph of some measurable mapping from $[0, \infty) \times S$ to A. The transition rate is given by a signed kernel q on S given K, assumed to satisfy $q(j|t,i,a) \geq 0$ if $j \neq i$ with $j, i \in S$, $q(S|t,i,a) \equiv 0$, and

$$q^*(i) := \sup_{t\geq 0, a\in A(t,i)} q(t,i,a) < \infty, \ \forall \ i \in S, \tag{2.1}$$

where $q(t,i,a) := -q(i|t,i,a) \geq 0$ for all $(t,i,a) \in K$. The running cost rate c is a measurable function on K. We shall consider the problem over a finite time duration. The terminal cost g is a function on S.

We briefly describe the construction of a CTMDP as in [14, 15]. Let $S_\Delta := S \bigcup \{\Delta\}$ (with some $\Delta \notin S$ being an isolated point), $\Omega^0 := (S \times (0,\infty))^\infty$ be the countable product. The canonical sample space Ω is the union of Ω^0 and all the sequences in the form of $(i_0, \theta_1, i_1, \ldots, \theta_k, i_k, \infty, \Delta, \infty, \ldots)$ for some $k \geq 0$ (accepting $\theta_0 := 0$). Let $\mathcal{F}$ be the Borel σ-algebra on Ω. For each $\omega \in \Omega$, introduce $T_0(\omega) := 0$, $T_{k+1}(\omega) := \theta_1 + \theta_2 + \ldots + \theta_{k+1}$, $X_k(\omega) := i_k$. In what follows, the argument ω is often omitted. Let $\mathcal{F}_t$ be the internal history of the marked point process $\{T_n, X_n\}$. Let $T_\infty := \lim_{k\to\infty} T_k$. The controlled process $\{\xi_t\}$ is defined by

$$\xi_t(\omega) := \sum_{k\geq 0} I_{\{T_k \leq t < T_{k+1}\}} i_k + \Delta I_{\{t \geq T_\infty\}}, \forall \ \ t \geq 0.$$

Here and below, I_E stands for the indicator function on any set E, and for notational convenience, we defined that $i \cdot 0 = 0$ and $i \cdot 1 = i$ for each $i \in S_\Delta$.

We do not intend to consider the controlled process after moment T_∞, and put

$$q(\cdot|t,\Delta,a_\Delta) :\equiv 0, \ r(t,\Delta,a_\Delta) :\equiv 0, \ A(t,\Delta) := \{a_\Delta\}, \ A_\Delta := A \cup \{a_\Delta\},$$

where $a_\Delta \notin A$ is an isolated point.

A (history-dependent) policy π is determined and often identified by a sequence of stochastic kernels $\{\pi^k, k \geq 0\}$ such that

$$\begin{aligned}\pi(da|\omega,t) &= I_{\{t=0\}}\pi^0(da|i_0,0) + \sum_{k\geq 0} I_{\{T_k < t \leq T_{k+1}\}} \pi^k(da|i_0,\theta_1,i_1,\ldots,\theta_k,i_k,t-T_k) \\ &\quad + I_{\{t\geq T_\infty\}} I_{\{a_\Delta\}}(da).\end{aligned}$$

A policy π is called Markov if, with slight abuse of notations, $\pi(da|\omega,t) = \pi(da|\xi_{t-},t)$, which is denoted by $\pi_t(da|\cdot)$, where $\xi_{t-} = \lim_{s\uparrow t}\xi_s$. A Markov policy $\pi_t(da|\cdot)$ is called deterministic Markov if there exists a measurable mapping f on $[0,\infty) \times S$ such that $\pi_t(da|i)$ is a Dirac measure concentrated at $f(t,i)$. A deterministic Markov policy will be denoted by the underlying measurable mapping f. We denote by Π the set of all policies, by Π_m^r the set of all Markov policies, and by Π_m^d the set of all deterministic Markov policies.

For each $\pi \in \Pi$, the random measure m^π defined by

$$m^\pi(j|\omega,t)dt := \int_A q(j \setminus \{\xi_{t-}\}|t,\xi_{t-},a)\pi(da|\omega,t)dt \tag{2.2}$$

is predictable, see [12]. For each $\pi \in \Pi$ and $i \in S$, let $\mathbb{P}_i^\pi$ be the probability on $(\Omega, \mathcal{F})$ such that $\mathbb{P}_i^\pi(\xi_0 = i) = 1$, and with respect to which, $m^\pi(j|\omega,t)dt$ is the dual predictable projection of the random measure $\sum_{n\geq 1}\delta_{(T_n,X_n)}(dt,dx)$ of the marked point process $\{T_n, X_n\}$ on $\mathcal{B}((0,\infty)\times S)$, see [12, 14] or Chapter 4 of [15] for more details. Let $\mathbb{E}_i^\pi$ be the expectation taken with respect to $\mathbb{P}_i^\pi$.

For the intuitive description, a CTMDP is a continuous-time Markov pure jump process whose local characteristics (transition intensity and post-jump distributions) are controlled. After the n-th jump, and a history of state and sojourn times $h_n = (i_0, \theta_1, \ldots, \theta_n, i_n)$ is observed with $\theta_n < \infty$, the conditional (joint) distribution of the next state and sojourn time is determined by $\int_A q(j|t_n +$

$t, i_n, a)\pi_n(da|h_n, t-t_n)e^{-\int_0^t \int_A q(s+t_n, i_n, a)\pi_n(da|h_n, s)ds}dt$, $j \neq i_n$, where t_n is the observed value of the n-th jump moment. In particular, the next sojourn time has the conditional distribution obeying a nonstationary exponential distribution, and a policy specifies the selection of an action at any time moment based on the observed history.

We consider the following optimal control problem over the finite time duration $T > 0$:

$$\text{Minimize over } \pi \in \Pi\text{: } \mathcal{V}(\pi, i) := \mathbb{E}_i^\pi \left[e^{\int_0^T \int_A c(t,\xi_t,a)\pi(da|\omega,t)dt + g(\xi_T)} \right]. \tag{2.3}$$

Conditions imposed in the next section guarantee that the above expectation and integral are well defined. For each $i \in S$, let

$$\mathcal{V}^*(i) = \inf_{\pi \in \Pi} \mathcal{V}(\pi, i).$$

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

Problem (2.3) is often said to be with a risk-sensitive criterion, as the exponential utility reflects that the decision maker is increasingly averse to the higher cost, see [11]. This is in contrast with a linear utility, which is called risk-neutral. In discrete-time, risk-sensitive Markov decision processes received increasing interest in the recent years, see [3, 4, 13, 17] for example. These works mainly consider infinite-horizon problems; in the discrete-time setup, problems on finite horizon can be readily solved using backward induction. See also [1], which considered a more general utility function.

The objective of this paper is to provide conditions that can be satisfied by unbounded transition and cost rates, under which, there exists a deterministic Markov optimal policy, and the optimality equation has a unique solution out of a certain class of functions. We present an example in the next section, demonstrating a natural application of CTMDPs to controlled queueing system, where the transition and cost rates are both unbounded and thus not covered by the previous literature.

3 Preliminaries

In this section, we impose a set of conditions allowing one to consider unbounded transition and cost rates, see Example 3.1 below, and present several preliminary statements, which will serve the proof of Theorem 4.1 below.

Condition 3.1. *There exist a $[1,\infty)$-valued function V defined on S and constants $\rho > 0$, $M > 1$ such that*

(a) $\sum_{j\in S} q(j|t,i,a)V(j) \leq \rho V(i)$ for each $(t,i,a) \in K$;

(b) $q^*(i) \leq MV(i)$ for all $i \in S$, where $q^*(i)$ is as in (2.1);

(c) $e^{2(1+T)|c(t,i,a)|} \leq MV(i)$ for each $(t,i,a) \in K$, and $e^{2(1+T)|g(i)|} \leq MV(i)$ for each $i \in S$.

Compared to the risk-neutral (linear utility) case, to ensure the performance to be finite in the risk-sensitive setup, it is necessary to impose more restrictive conditions on the growth of the cost rate. Part (c) of Condition 3.1 is motivated by part (b) of the next lemma and the Jensen inequality, see the proof of Lemma 3.1 below.

Lemma 3.1. *Suppose Condition 3.1 is satisfied. For each $\pi \in \Pi$, the following assertions hold.*

(a) $\mathbb{P}_i^\pi(T_\infty = \infty) = 1$ for each $i \in S$.

(b) $\mathbb{E}_i^\pi[V(\xi_t)] \leq e^{\rho t}V(i)$, for each $t \geq 0$ and $i \in S$.

(c) $V(\pi, i) \leq Me^{T\rho}V(i)$ for all $i \in S$ and $\pi \in \Pi$.

Proof. Parts (a) and (b) are known, see e.g., [8, 19, 20]. We next verify part (c). By part (a), for $\mathbb{P}_i^\pi$-almost all $\omega \in \Omega$, there are finitely many values taken by in $\{\xi_t(\omega)\}$ over $[0, T]$. For such $\omega \in \Omega$, by Condition 3.1(c), we legitimately write

$$\int_0^T \int_A c(t, \xi_t, a)\pi(da|\omega, t)dt + g(\xi_T) = \int_{(0,T]} \int_A \tilde{c}(t, \xi_t, a)\pi(da|\omega, t)\mu(dt),$$

where $\mu(dt) = I_{[0,T)}(t)dt + \delta_T(dt)$, with $\delta_T(dt)$ being the Dirac measure concentrated on $\{T\}$, and $\tilde{c}(t, i, a) := c(t, i, a)I_{[0,T)}(t) + g(i)I_{\{T\}}(t)$ for each $(t, i, a) \in K$. Now,

$$\begin{aligned}
& \mathbb{E}_i^\pi \left[e^{\int_0^T \int_A c(t,\xi_t,a)\pi(da|\omega,t)dt + g(\xi_T)} \right] = \mathbb{E}_i^\pi \left[e^{\int_{[0,T]} \int_A (1+T)\tilde{c}(t,\xi_t,a)\pi(da|\omega,t)\frac{\mu(dt)}{T+1}} \right] \\
\leq \; & \mathbb{E}_i^\pi \left[\frac{1}{1+T} \int_{[0,T]} e^{(1+T)\int_A |\tilde{c}(t,\xi_t,a)|\pi(da|\omega,t)} \mu(dt) \right] \leq \frac{M}{1+T} \mathbb{E}_i^\pi \left[\int_0^T V(\xi_t)dt + V(\xi_T) \right] \\
\leq \; & Me^{\rho T}V(i), \qquad (3.1)
\end{aligned}$$

where the first inequality is by the Jensen inequality, the second inequality is by Condition 3.1(c), and the last inequality is by part (b). □

Part (a) of the previous lemma asserts that under the imposed conditions therein, the controlled process is nonexplosive under each policy. This fact is used in the proof of Theorem 3.1 below, see the first paragraph therein as well as (3.7).

Condition 3.2. *There exist a $[1, \infty)$-valued function V_1 defined on S, and constants $\rho_1 > 0$, $M_1 > 0$ such that*

(a) $\sum_{j \in S} V_1^2(j)q(j|t, i, a) \leq \rho_1 V_1^2(i)$ for each $(t, i, a) \in K$;

(b) $V^2(i) \leq M_1 V_1(i)$ for all $i \in S$, with the function V as the Condition 3.1.

The role of this condition is seen in the proof of Theorem 3.1, where the Cauchy-Schwarz inequality is used, see (3.4) therein. Conditions 3.1 and 3.2 guarantee the growth of the value function and its derivative to be suitably bounded by the weight functions V and V_1, and it is out of this class of functions that we show the Feynman-Kac formula applies. The previous works [5, 21] only showed that the Feynman-Kac formula is applicable to a class of bounded functions, and so confined themselves to the class of bounded cost rates, which excludes some potentially interesting applications. Let us formulate such an example, which are with unbounded transition and cost rates and satisfy Conditions 3.1 and 3.2.

Example 3.1. *Consider a controlled $M/M/\infty$ queueing system, where the common service rate a of each server can be tuned from a finite interval $[\underline{\mu}, \overline{\mu}] \subseteq [0, \infty]$. Let the arrival rate be denoted by $\lambda > 0$. The holding cost is $C_1 i$ given the current number of jobs in the system being $i \geq 0$, where $C_1 > 0$ is a constant, and maintaining a service rate at μ costs μ per unit time. A terminal reward of $C_2 i$ is received if there are i jobs remaining in the system at the end of the horizon $[0, T]$, where $C_2 \in (-\infty, \infty)$ is a constant. The decision maker aims at the optimal control of the service rate to minimize the expected exponential utility of the total cost over the horizon $[0, T]$.*

This problem can be formulated as a CTMDP with the following primitives. The state space is $S = \{0, 1, \dots\}$, the action space is $[\underline{\mu}, \overline{\mu}] \equiv A(t, i)$. The transition rate is given by $q(i + 1|t, i, a) \equiv \lambda$,

$q(i-1|t,i,a)=ai$ if $i\geq 1$, $q(t,i,a)=\lambda+ai$ if $i>0$, and $q(t,0,a)=\lambda$. The running cost rate is given by $c(t,i,a)=C_1i+a$, and the terminal cost is given by $g(i)=-C_2i$.

Observe the following. Let $d>0$ be a fixed constant. Let $\rho(d):=e^{d+1}\lambda$. Then for each constant $\rho\geq\rho(d)$, $\sum_{j\in S}q(j|t,i,a)e^{dj}=e^{d(i+1)}\lambda+e^{d(i-1)}a-(\lambda+a)e^{di}\leq\rho e^{di}$ for each $i\geq 1$, and $\sum_{j\in S}q(j|t,0,a)e^{dj}=\lambda e^{d}-\lambda\leq\rho$. Therefore, for the verification of Condition 3.1, one can take $M=e^{2(1+T)\overline{\mu}}+\overline{\mu}+\lambda$, $V(i)=e^{d_1i}$ with $d_1=2(1+T)(C_1+|C_2|)$, $\rho=\rho(d_1)$. For the verification of Condition 3.2, one can take $M_1=1$, and $V_1(i)=e^{d_2i}$ with $d_2=2d_1$, and $\rho_1=\rho(d_2)$.

Let us introduce some additional notations, which will be needed in the next statement. In particular, it formalizes what we mean in the Introduction by "a large enough class of functions" to which, the Feynman-Kac formula applies. Let X be a Borel space. For each measurable function ψ on $[0,T]\times X$, if $\psi(\cdot,x)$ is absolutely continuous on $[0,T]$, then we put ψ' a measurable function on $[0,T]\times X$ such that $\psi(t,x)-\psi(0,x)=\int_0^t\psi'(s,x)ds$ for each $x\in X$ and $t\in[0,T]$. Consider the functions V and V_1 as in Conditions 3.1 and 3.2. A function φ on $[0,T]\times S$ is called V-bounded if the V-weighted norm of φ, $\|\varphi\|_V:=\sup_{(t,i)\in[0,T]\times S}\frac{|\varphi(t,i)|}{V(i)}$, is finite. Let $C^1_{V,V_1}([0,T]\times S)$ be the collection of V-bounded functions φ on $[0,T]\times S$ such that $\varphi(\cdot,i)$ is absolutely continuous on $[0,T]$ for each $i\in S$, which admits some V_1-bounded φ'.

Theorem 3.1. *Suppose Conditions 3.1 and 3.2 are satisfied. Then, for each $i\in S$, $\pi\in\Pi$ and $\varphi\in C^1_{V,V_1}([0,T]\times S)$,*

$$\begin{aligned}&\mathbb{E}_i^\pi\left[\int_0^T\left(\psi'(\omega,t,\xi_t)+\sum_{j\in S}\psi(\omega,t,j)\int_A q(j|t,\xi_t,a)\pi(da|\omega,t)\right)dt\right]\\ =\ &\mathbb{E}_i^\pi\left[\psi(\omega,T,\xi_T)\right]-\varphi(0,i),\end{aligned}$$

where outside a $\mathbb{P}_i^\pi$-null set, say $\Omega\setminus\Omega'$, $T_\infty=\infty$,

$$\psi(\omega,t,j)=e^{\int_0^t\int_A c(v,\xi_v,a)\pi(da|\omega,v)dv}\varphi(t,j),\ \forall\ t\in[0,T],\ j\in S,$$

$\psi(\omega,\cdot,j)$ is absolutely continuous on $[0,T]$ so that we can take

$$\begin{aligned}\psi'(\omega,t,j)\ =\ &\int_A c(t,\xi_t,a)\pi(da|\omega,t)e^{\int_0^t\int_A c(v,\xi_v,a)\pi(da|\omega,v)dv}\varphi(t,j)\\ &+e^{\int_0^t\int_A c(v,\xi_v,a)\pi(da|\omega,v)dv}\varphi'(t,j),\end{aligned}\tag{3.2}$$

for each $\omega\in\Omega'$ and $j\in S$.

Proof. According to Lemma 3.1(a), we concentrate on Ω' on which $T_\infty=\infty$, and hence (3.2) holds. Since $\varphi\in C^1_{V,V_1}([0,T]\times S)$, we have $|\varphi(t,i)|\leq\|\varphi\|_VV(i)$ for all $(t,i)\in[0,T]\times S$, which, together with the relation $(1+T)|c(v,i,a)|\leq MV(i)$ (by Condition 3.1(c)), leads to

$$\begin{aligned}&\left|\psi'(\omega,t,\xi_t)\right|\\ \leq\ &\frac{M}{1+T}V(\xi_t)e^{\int_0^t\int_A|c(v,\xi_v,a)|\pi(da|\omega,v)dv}\|\varphi\|_VV(\xi_t)+\|\varphi'\|_{V_1}e^{\int_0^t\int_A|c(v,\xi_v,a)|\pi(da|\omega,v)dv}V_1(\xi_t),\\ \leq\ &\frac{\|\varphi\|_V+\|\varphi'\|_{V_1}}{1+T}(1+T+MM_1)e^{\int_0^t\int_A|c(v,\xi_v,a)|\pi(da|\omega,v)dv}V_1(\xi_t).\end{aligned}\tag{3.3}$$

By the Cauchy-Schwarz inequality,

$$\mathbb{E}_i^\pi\left[e^{\int_0^t\int_A|c(v,\xi_v,a)|\pi(da|\omega,v)dv}V_1(\xi_t)\right]\leq\sqrt{\mathbb{E}_i^\pi\left[e^{2\int_0^t\int_A|c(v,\xi_v,a)|\pi(da|\omega,v)dv}\right]\mathbb{E}_i^\pi\left[V_1^2(\xi_t)\right]}$$

$$
\begin{aligned}
&\leq \mathbb{E}_i^\pi\left[e^{2\int_0^t\int_A|c(v,\xi_v,a)|\pi(da|\omega,v)dv}\right]\mathbb{E}_i^\pi\left[V_1^2(\xi_t)\right] \leq Me^{T\rho}V(i)\mathbb{E}_i^\pi\left[V_1^2(\xi_t)\right]\\
&\leq Me^{T\rho}V(i)e^{\rho_1 T}V_1^2(i),\ t\in[0,T], \qquad (3.4)
\end{aligned}
$$

where the second to the last inequality is obtained by a similar argument to the one for (3.1), and the last inequality is by Lemma 3.1(b). Now it follows from (3.3) that

$$
\mathbb{E}_i^\pi\left[\int_0^T|\psi'(\omega,t,\xi_t)|dt\right]<\infty. \tag{3.5}
$$

On the other hand, by Conditions 3.1 and 3.2, we have

$$
\begin{aligned}
&\sum_{j\in S}e^{\int_0^t\int_A|c(v,\xi_v,a)|\pi(da|\omega,v)dv}|\varphi(t,j)|\left|\int_A q(j|t,\xi_t,a)\pi(da|\omega,t)\right|\\
\leq\ & \|\varphi\|_V\left(\rho V(\xi_t)+2MV^2(\xi_t)\right)e^{\int_0^t\int_A|c(v,\xi_v,a)|\pi(da|\omega,v)dv}\\
\leq\ & \|\varphi\|_V M_1(\rho+2M)e^{\int_0^t\int_A|c(v,\xi_v,a)|\pi(da|\omega,v)dv}V_1(\xi_t).
\end{aligned}
$$

Now it follows from (3.4) that

$$
\int_0^T\sum_{j\in S}\mathbb{E}_i^\pi\left[\left|\int_A q(j|t,\xi_t,a)\pi(da|\omega,t)\right||\psi(\omega,t,j)|\right]dt<\infty. \tag{3.6}
$$

For each $0\leq s\leq T$,

$$
\psi(\omega,T,\xi_T)=\psi(\omega,0,\xi_0)+\int_0^T\psi'(\omega,t,\xi_t)dt+\sum_{n\geq 1}\int_{(0,T]}\Delta\psi(\omega,t,\xi_t)\delta_{T_n}(dt) \tag{3.7}
$$

with $\Delta\psi(\omega,t,\xi_t):=\psi(\omega,t,\xi_t)-\psi(\omega,t-,\xi_{t-})$. (Recall that the function $\psi(\omega,t,j)$ is absolutely continuous in t over finite interval, and for each fixed $\omega\in\Omega'$ with Ω' being defined in the beginning of this proof, $\xi_t(\omega)$ is piecewise constant in $t\in[0,T]$, and assumes finitely many values over that interval.) By (3.5) and (3.6), we take legitimately the expectation on the both sides of the previous equality, and obtain

$$
\begin{aligned}
&\mathbb{E}_i^\pi\left[\psi(\omega,T,\xi_T)\right]=\mathbb{E}_i^\pi\left[\psi(\omega,0,\xi_0)\right]+\mathbb{E}_i^\pi\left[\int_0^T\psi'(\omega,t,\xi_t)dt\right]\\
&+\mathbb{E}_i^\pi\left[\sum_{n\geq 1}\int_{(0,T]}\Delta\psi(\omega,t,\xi_t)\delta_{T_n}(dt)\right]\\
=\ &\varphi(0,i)+\mathbb{E}_i^\pi\left[\int_0^T\psi'(\omega,t,\xi_t)dt\right]\\
&+\mathbb{E}_i^\pi\left[\sum_{j\in S}\int_{(0,T]}(\psi(\omega,t,j)-\psi(\omega,t,\xi_{t-}))m^\pi(j|\omega,t)dt\right]\\
=\ &\varphi(0,i)+\mathbb{E}_i^\pi\left[\int_0^T\psi'(\omega,t,\xi_t)dt\right]\\
&+\mathbb{E}_i^\pi\left[\sum_{j\in S}\int_0^T\int_A\psi(\omega,t,j)q(j|t,\xi_{t-},a)\pi(da|\omega,t)dt\right],
\end{aligned}
$$

where the last equality holds because the random measure m^π defined by (2.2) is the dual predictable projection of the random measure $\sum_{n\geq 1}\delta_{(T_n,X_n)}(dt,dx)$ on $\mathcal{B}((0,\infty)\times S)$ under $\mathbb{P}_i^\pi$, see p.131 of [15]. The statement is proved. □

The above Feynman-Kac formula in the above theorem was justified in [21], see Theorem 3.1 therein, when π is a Markov policy, and φ is assumed to be bounded.

The next statement provides a verification theorem, which was known in [18] when the transition rate is bounded.

Corollary 3.1. *Suppose Conditions 3.1 and 3.2 are satisfied. If there exists $\varphi \in C^1_{V,V_1}([0,T]\times S)$ and a deterministic Markov policy $f\in \Pi^d_m$ such that*

$$\begin{aligned}\varphi(s,i)-e^{g(i)} &= \int_s^T \inf_{a\in A(t,i)}\left\{c(t,i,a)\varphi(t,i)+\sum_{j\in S}\varphi(t,j)q(j|t,i,a)\right\}dt\\ &= \int_s^T \left\{c(t,i,f(t,i))\varphi(t,i)+\sum_{j\in S}\varphi(t,j)q(j|t,i,f(t,i))\right\}dt,\\ & \quad s\in[0,T],\ \ i\in S, \qquad (3.8)\end{aligned}$$

then

$$\mathcal{V}(f,i)=\varphi(0,i)=\mathcal{V}^*(i),\ \forall\ i\in S. \qquad (3.9)$$

Proof. Concentrate on Ω' as in the proof of the previous theorem. It holds for almost all $t\in[0,T]$ that

$$\begin{aligned}0=\varphi'(t,\xi_t)&+\inf_{a\in A(t,\xi_t)}\left\{c(t,\xi_t,a)\varphi(t,\xi_t)+\sum_{j\in S}\varphi(t,j)q(j|t,\xi_t,a)\right\}\\ =&\ \ \varphi'(t,\xi_t)+c(t,\xi_t,f(t,\xi_t))\varphi(t,\xi_t)+\sum_{j\in S}\varphi(t,j)q(j|t,\xi_t,f(t,\xi_t))\\ \leq&\ \ \varphi'(t,\xi_t)+\int_A\left\{c(t,\xi_t,a)\varphi(t,\xi_t)+\sum_{j\in S}\varphi(t,j)q(j|t,\xi_t,a)\right\}\pi(da|\omega,t).\end{aligned}$$

Now by applying Theorem 3.1 to the deterministic Markov policy f and an arbitrarily fixed $\pi\in\Pi$, we see

$$\begin{aligned}\mathcal{V}(\pi,i)-\varphi(0,i)&=\mathbb{E}_i^\pi\left[e^{\int_0^T\int_A c(v,\xi_v,a)\pi(da|\omega,v)dv}\varphi(T,\xi_T)\right]-\varphi(0,i)\\ =\ \ &\mathbb{E}_i^\pi\left[\int_0^T e^{\int_0^t\int_A c(v,\xi_v,a)\pi(da|\omega,v)dv}\int_A(c(t,\xi_t,a)\varphi(t,\xi_t)+\varphi'(t,\xi_t)+\sum_{j\in S}\varphi(t,j)q(j|t,\xi_t,a))\pi(da|\omega,t)\right]\\ \geq\ \ &0,\end{aligned}$$

where the first equality holds because $\varphi(T,i)=e^{g(i)}$, see (3.8); similarly, replacing f for π in the equalities in the above, $\mathcal{V}(f,i)-\varphi(0,i)=0$. Consequently, $\mathcal{V}(f,i)=\varphi(0,i)\leq\mathcal{V}(\pi,i)$ for each $i\in S$. Since π was arbitrarily fixed, $\mathcal{V}(f,i)=\varphi(0,i)=\mathcal{V}^*(i)$, as required. □

According to the previous statement, (3.8) is called the optimality equation, and the policy f in (3.9) is optimal.

The next statement was basically obtained in Theorem 2.1 in [5], see also [21].

Proposition 3.1. *Suppose that the transition and cost rates are bounded, i.e.,*

$$\sup_{i\in S} q^*(i) < \infty, \quad \sup_{(t,i,a)\in K} |c(t,i,a)| < \infty, \quad \sup_{i\in S} |g(i)| < \infty.$$

If for each $i \in S$ and $t \in [0,T]$, $A(t,i)$ is compact, $c(t,i,a)$ is lower semicontinuous in $a \in A(t,i)$, and $q(j|t,i,a)$ is continuous in $a \in A(t,i)$, then there exists a unique φ in $C^1_{1,1}([0,T] \times S)$ and some $f \in \Pi^d_m$ satisfying (3.8) and (3.9).

The main objective in this paper is to relax the boundedness requirements in the previous statement.

4 Optimality result

We impose the following condition, which guarantees the existence of an optimal policy.

Condition 4.1. *(a) For each $(t,i) \in [0,T] \times S$, $A(t,i)$ is compact.*

(b) For each $t \in [0,T], i,j \in S$, the function $q(j|t,i,a)$ is continuous in $a \in A(t,i)$.

(c) For each $(t,i) \in [0,T] \times S$, the function $c(t,i,a)$ is lower semicontinuous in $a \in A(t,i)$, and the function $\sum_{j\in S} V(j)q(j|t,i,a)$ is continuous in $a \in A(t,i)$, with V as in Condition 3.1.

Under Conditions 3.1 and 4.1(b) and (c), the function $\sum_{j\in S} q(j|t,i,a)u(t,j)$ is continuous in $a \in A(t,i)$, for every fixed $(t,i) \in [0,T] \times S$ and V-bounded measurable function u on $[0,T] \times S$, see the proof of Lemma 8.3.7(a) in [10]. This fact will be used in the proof of the next statement.

Also note that Condition 4.1 is satisfied by Example 3.1.

The main optimality result is the following one.

Theorem 4.1. *Suppose Conditions 3.1, 3.2 and 4.1 are satisfied. Then there exists a unique φ in $C^1_{V,V_1}([0,T]\times S)$ and some $f \in \Pi^d_m$ satisfying (3.8) and (3.9). In particular, there exists a deterministic Markov optimal policy.*

Proof. The statement would follow from Corollary 3.1, once we showed the existence of some $\varphi \in C^1_{V,V_1}([0,T] \times S)$ satisfying (3.8). We verify this fact following a similar reasoning as in [7] dealing with a risk-neutral CTMDP problem, which was also adopted in [21], dealing with a model with a bounded cost rate. Namely, we shall obtain the desired solution φ as a limit point of an equicontinuous family $\{\varphi_n\}$ of functions, which in turn are obtained from a sequence of CTMDP models with bounded transition and cost rates. The denumerable state space serves to prove the equicontinuity of the family $\{\varphi_n\}$. The details are as follows.

For each integer $n \geq 1$, let $S_n := \{i \in S : \ V(i) \leq n\}$. Without loss of generality, assume for each $n \geq 1$, $S_n \neq \emptyset$. For each $i \in S$ and $t \in [0,\infty)$, let $A_n(t,i) := A(t,i)$. For each $(t,i,a) \in K_n := K$, define

$$q_n(j|t,i,a) := q(j|t,i,a)I_{S_n}(i), \ \forall \ j \in S, \ c_n(t,i,a) := c(t,i,a)I_{S_n}(i), \ g_n(i) := g(i)I_{S_n}(i).$$

We consider the resulting sequence of CTMDP models $\mathcal{M}_n := \{S, A_n(t,i), c_n, g_n, q_n\}$.

Note that the models $\{\mathcal{M}_n\}$ are all with bounded transition and cost rates, and so Proposition 3.1 implies, for each $n \geq 1$, the existence of a unique φ_n in $C^1_{1,1}([0,T] \times S)$ and some $f_n \in \Pi^d_m$ satisfying

$$\varphi_n(s,i) - e^{g_n(i)} \quad = \quad \int_s^T \inf_{a\in A(t,i)} \left\{ c_n(t,i,a)\varphi_n(t,i) + \sum_{j\in S} \varphi_n(t,j)q_n(j|t,i,a) \right\} dt$$

$$
\begin{aligned}
&= \int_s^T \left\{ c_n(t,i,f_n(t,i))\varphi_n(t,i) + \sum_{j\in S} \varphi_n(t,j) q_n(j|t,i,f_n(t,i)) \right\} dt, \\
&\quad s \in [0,T],\ i \in S. \qquad (4.1)
\end{aligned}
$$

Let $n \geq 1$ be fixed. For each $s \in [0,T]$, consider the s-shifted model

$$\mathcal{M}_n^{(s)} := \left\{ S, A_n^{(s)}(t,i), q_n^{(s)}, c_n^{(s)}, g_n \right\}$$

with $A_n^{(s)}(t,i) := A_n(t+s,i)$, $q_n^{(s)}(\cdot|t,i,a) := q_n(\cdot|s+t,i,a)$ and $c_n^{(s)}(t,i,a) := c_n(t+s,i,a)$. Then Condition 3.1 is clearly satisfied by $\mathcal{M}_n^{(s)}$, so that one can apply the reasoning in the proof of Lemma 3.1(c) and deduce

$$\mathrm{E}_i^{f_n^{(s)}} \left[e^{\int_0^{T-s} |c_n^{(s)}(t,\xi_t,f_n^{(s)}(t,\xi_t))| dt + |g_n(\xi_{T-s})|} \right] \leq M e^{T\rho} V(i)$$

where $\mathrm{E}_i^{f_n^{(s)}}$ denotes the expectation in the $\mathcal{M}_n^{(s)}$ model under the shifted policy $f_n^{(s)}(t,i) := f_n(t+s,i)$. On the other hand, according to the uniqueness of the solution to (4.1) in $C^1_{1,1}([0,T]\times S)$ and the discussions at the end of Section 3 of [6] after Theorem 3.2 therein,

$$\mathrm{E}_i^{f_n^{(s)}} \left[e^{\int_0^{T-s} c_n^{(s)}(t,\xi_t,f_n^{(s)}(t,\xi_t)) dt + g_n(\xi_{T-s})} \right] = \varphi_n(s,i).$$

(The cost rate and the terminal cost were assumed to be nonnegative in [6], but the results obtained there apply because $\mathcal{M}_n^{(s)}$ has bounded transition and cost rates, which can be reduced to the nonnegative case after one add to the cost rate and the terminal cost a large enough constant.) Thus, we obtain the bound

$$|\varphi_n(t,i)| \leq M e^{T\rho} V(i),\ \forall\ n \geq 1, (t,i) \in [0,T]\times S. \qquad (4.2)$$

Next, we show that $\{\varphi_n, n \geq 1\}$ is an equicontinuous family of functions on $[0,T]\times S$, as follows. Let

$$H_n(t,i) := \inf_{a\in A_n(t,i)} \left\{ c_n(t,i,a)\varphi_n(t,i) + \sum_{j\in S} \varphi_n(t,j) q_n(j|t,i,a) \right\},\ \forall\ (t,i) \in [0,T]\times S.$$

Then, from Condition 3.1 and (4.2), we see

$$
\begin{aligned}
|H_n(t,i)| &\leq \sup_{a\in A_n(t,i)} \left\{ |c_n(t,i,a)\varphi_n(t,i)| + \sum_{j\in S} |\varphi_n(t,j)||q_n(j|t,i,a)| \right\} \\
&\leq \sup_{a\in A_n(t,i)} \left\{ MV(i)Me^{T\rho}V(i) + Me^{T\rho} \sum_{j\in S} |q(j|t,i,a)|V(j) \right\} \\
&\leq e^{T\rho}(M^2V^2(i) + \rho MV(i) + 2M|q(i|t,i,a)|V(i)) \\
&\leq Me^{T\rho}M_1(3M^2+\rho)V_1(i) =: L(i),\ \forall\ (t,i) \in [0,T]\times S. \qquad (4.3)
\end{aligned}
$$

(Recall that $M > 1$.)

Now, fix arbitrarily some $(s_0,i_0) \in [0,T]\times S$ and $\varepsilon > 0$, and take $\delta := \min\{\frac{\varepsilon}{L(i_0)}, \frac{1}{2}\}$. Then, for every (s,i) in the open neighborhood $\{(s,i) \in [0,T]\times S:\ |s-s_0| < \delta, |i-i_0| < \delta\}$, we have $i = i_0$, and

$$|\varphi_n(s,i) - \varphi_n(s_0,i_0)| = |\varphi_n(s,i_0) - \varphi_n(s_0,i_0)| = \left| \int_s^T H_n(t,i_0)dt - \int_{s_0}^T H_n(t,i_0)dt \right|$$

$$\leq \quad L(i_0)|s - s_0| < \varepsilon, \ \forall \ n \geq 1.$$

Hence, $\{\varphi_n, n \geq 1\}$ is equicontinuous at (s_0, i_0), which, together with the arbitrariness of $(s_0, i_0) \in [0,T] \times S$, yields that $\{\varphi_n, n \geq 1\}$ is equicontinuous on $[0,T] \times S$. By Arzela-Ascoli theorem, see, e.g., p.96 of [9], there exist a subsequence $\{\varphi_{n_k}, k \geq 1\}$ of $\{\varphi_n, n \geq 1\}$ and a continuous function φ on $[0,T] \times S$ such that

$$\lim_{k\to\infty} \varphi_{n_k}(s,i) = \varphi(s,i), \text{ and } |\varphi(s,i)| \leq Me^{T\rho}V(i) \ \ \forall \ (s,i) \in [0,T] \times S, \tag{4.4}$$

where the last inequality is by (4.2).

Let

$$H(t,i) := \inf_{a\in A(t,i)} \left\{ c(t,i,a)\varphi(t,i) + \sum_{j\in S} \varphi(t,j)q(j|t,i,a) \right\}, \forall \ (t,i) \in [0,T] \times S.$$

We next verify that $\lim_{k\to\infty} H_{n_k}(t,i) = H(t,i)$ for each $(t,i) \in [0,T] \times S$, as follows. Let $(t,i) \in [0,T] \times S$ be arbitrarily fixed. Since $q_{n_k}(j|t,i,a) \to q(j|t,i,a)$ for all $j \in S$ and $a \in A(t,i)$ as $k \to \infty$, by virtue of Lemma 8.3.7 in [10] and (4.2), we have

$$\begin{aligned} \limsup_{k\to\infty} H_{n_k}(t,i) \ &\leq \ \limsup_{k\to\infty} \left\{ c_{n_k}(t,i,a)\varphi_{n_k}(t,i) + \sum_{j\in S} \varphi_{n_k}(t,j)q_{n_k}(j|t,i,a) \right\} \\ &\leq \ c(t,i,a)\varphi(t,i) + \sum_{j\in S} \varphi(t,j)q(j|t,i,a), \ \forall \ a \in A(t,i), \end{aligned}$$

so that

$$\limsup_{k\to\infty} H_{n_k}(t,i) \ \leq \ \inf_{a\in A(t,i)} \left\{ c(t,i,a)\varphi(t,i) + \sum_{j\in S} \varphi(t,j)q(j|t,i,a) \right\}. \tag{4.5}$$

According to the fact mentioned below Condition 4.1, there exists a sequence of policies $\{f_{n_k}\} \subseteq \Pi_m^d$ such that

$$\begin{aligned} H_{n_k}(t,i) \ &= \ \inf_{a\in A(t,i)} \left\{ c_{n_k}(t,i,a)\varphi_{n_k}(s,i) + \sum_{j\in S} \varphi_{n_k}(t,j)q_{n_k}(j|t,i,a) \right\} \\ &= \ c(t,i,f_{n_k}(t,i))\varphi_{n_k}(t,i) + \sum_{j\in S} \varphi_{n_k}(t,j)q_{n_k}(j|t,i,f_{n_k}(t,i)). \end{aligned}$$

Since $A(t,i)$ is compact, by taking subsequences if necessary, we can assume without loss of generality that $\liminf_{k\to\infty} H_{n_k}(t,i) = \lim_{k\to\infty} H_{n_k}(t,i)$ and for some $a \in A(t,i)$, $f_{n_k}(t,i) \to a$ as $k \to \infty$. By the virtue of Lemma 8.3.7 in [10], we have

$$\liminf_{k\to\infty} H_{n_k}(t,i) = \liminf_{k\to\infty} \left\{ c(t,i,f_{n_k}(t,i))\varphi_{n_k}(t,i) + \sum_{j\in S} \varphi_{n_k}(t,j)q_{n_k}(j|t,i,f_{n_k}(t,i)) \right\}$$

$$\geq \ c(s,i,a)\varphi(t,i) + \sum_{j\in S} \varphi(t,j)q(j|t,i,a) \geq \inf_{a\in A(t,i)} \left\{ c(t,i,a)\varphi(t,i) + \sum_{j\in S} \varphi(t,j)q(j|t,i,a) \right\}.$$

(Recall Condition 4.1.) This, together with (4.5), implies that $\lim_{k\to\infty} H_{n_k}(s,i) = H(s,i)$. Since $(s,i) \in [0,T]\times S$ was arbitrarily fixed, we see from (4.1), (4.3) and (4.4) that φ satisfies (3.8). The same argument as in (4.3) leads to

$$|\varphi'(t,i)| = |H(t,i)| \le Me^{T\rho}M_1(3M^2+\rho)V_1(i),\ \forall\ (t,i)\in[0,T]\times S.$$

Therefore, we see that $\varphi \in C^1_{V,V_1}([0,T]\times S)$. The required deterministic Markov policy f exists because of the fact mentioned below Condition 4.1, a measurable selection theorem, see Proposition D.5 of [9].

Finally, we verify the uniqueness part. Let $\varphi \in C^1_{V,V_1}([0,T]\times S)$ be an arbitrarily fixed solution to (3.8). (The above reasoning shows that there exists at least one.) Let $s\in[0,T]$ be fixed, and consider the s-shifted model $\mathcal{M}^{(s)} = \{S, A^{(s)}(t,i), q^{(s)}, c^{(s)}, g\}$, which is defined as for the $\mathcal{M}^{(s)}_n$ model with n being omitted everywhere. Let

$$V^{(s)}(i) := \inf_{\pi\in\Pi} \mathrm{E}^{\pi}_i\left[e^{\int_0^{T-s}\int_A c^{(s)}(t,\xi_t,a)\pi(da|\omega,t)dt+g(\xi_{T-s})}\right]$$

with E^{π}_i signifying the expectation in the s-shifted model. Then the function $\varphi^{(s)} \in C^1_{V,V_1}([0,T-s]\times S)$ defined by $\varphi^{(s)}(\tau,i) := \varphi(\tau+s,i)$ for each $(\tau,i)\in[0,T-s]\times S$ satisfies

$$\begin{aligned}\varphi^{(s)}(\tau,i) - e^{g(i)} &= \int_\tau^{T-s} \inf_{a\in A^{(s)}(t,i)}\left\{c^{(s)}(t,i,a)\varphi^{(s)}(t,i) + \sum_{j\in S}\varphi^{(s)}(t,j)q^{(s)}(j|t,i,a)\right\}dt \\ &= \int_\tau^{T-s}\left\{c^{(s)}(t,i,f^{(s)}(t,i))\varphi^{(s)}(t,i) + \sum_{j\in S}\varphi^{(s)}(t,j)q^{(s)}(j|t,i,f^{(s)}(t,i))\right\}dt, \\ &\quad \tau\in[0,T-s],\ i\in S,\end{aligned}$$

for some deterministic Markov policy $f^{(s)}$. By applying Corollary 3.1 to the s-shifted model $\mathcal{M}^{(s)}$, we see $\varphi^{(s)}(0,i) = V^{(s)}(i)$, and thus $\varphi(s,i) = V^{(s)}(i)$ for each $i\in S$. Since $s\in[0,T]$ was arbitrarily fixed, it follows that φ is the unique solution to (3.8) out of $\varphi\in C^1_{V,V_1}([0,T]\times S)$. The proof is completed. □

5 Conclusion

In this paper, we considered a risk-sensitive CTMDP problem in a denumerable state space over a finite time duration. Under conditions that can be satisfied by unbounded transition and cost rates, the optimality equation was shown to have a unique solution out of a class of functions, to which Feynman-Kac formula was shown to be applicable. The results obtained in this paper can be viewed as a response to the remark in Section 7 of [21], and complemented the relevant results in [6].

Acknowledgement

This work is partially supported by Natural Science Foundation of Guangdong Province (Grant No.2014A030313438), Zhujiang New Star (Grant No. 201506010056), Guangdong Province outstanding young teacher training plan (Grant No. YQ2015050).

Ethical statements

- The paper has not been published in whole or in part elsewhere;

- The manuscript is not currently being considered for publication in another journal;
- All authors have been personally and actively involved in substantive work leading to the manuscript, and will hold themselves jointly and individually responsible for its content.
- There is no potential conflicts of interest.
- Research do not have Human Participants and/or Animals.
- The funding bodies have no objections against publishing this paper.

References

[1] Bäuerle, N. and Rieder, U. (2014). More risk-sensitive Markov decision processes. *Math. Oper. Res.* **39**, 105–120.

[2] Bäuerle, N. and Popp, A. (2018). Risk-sensitive stopping problems for continuous-time Markov chains. *Stochastics* **90**, 411-431.

[3] Cavazos-Cadena, R. and Montes-de-Oca, R. (2000). Optimal stationary policies in risk-sensitive dynamic programs with finite state space and nonnegative rewards. *Applications Mathematicae* **27**, 167-185.

[4] Cavazos-Cadena, R. and Montes-de-Oca, R. (2000). Nearly optimal policies in risk-sensitive positive dynamic programming on discrete spaces. *Math. Meth. Oper. Res.* **52**, 133-167.

[5] Ghosh, M. and Saha, S. (2014). Risk-sensitive control of continuous time Markov chains. *Stochastics* **86**, 655–675

[6] Guo, X. and Zhang, Y. (2018) On risk-sensitive piecewise deterministic Markov decision processes. *Appl. Math. Optim.*, in press, https://doi.org/10.1007/s00245-018-9485-x

[7] Guo, X.P., Huang, X. and Huang, Y. (2015). Finite-horizon optimality for continuous-time Markov decision processes with unbounded transition rates. *Adv. in Appl. Probab.* **47**, 1064–1087.

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

[9] Hernández-Lerma, O. and Lasserre, J. (1996). *Discrete-Time Markov Control Processes.* Springer-Verlag, New York.

[10] Hernández-Lerma, O. and Lasserre, J. (1999). *Further Topics on Discrete-Time Markov Control Processes.* Springer-Verlag, New York.

[11] Howard, R. and Matheson, J. (1972). Risk-sensitive Markov decision proceses. *Manag. Sci.* **18**, 356–369.

[12] Jacod, J. (1975). Multivariate point processes: Predictable projection, Radon-Nicodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie und verwandte Gebiete* **31**, 235–253.

[13] Jaśkiewicz, A. (2008). A note on negative dynamic programming for risk-sensitive control. *Oper. Res. Lett.* **36**, 531-534.

[14] Kitaev, M. (1986). Semi-Markov and jump Markov controlled models: average cost criterion. *Theory. Probab. Appl.* **30**, 272–288.

[15] Kitaev, M. and Rykov, V. (1995). *Controlled Queueing Systems.* CRC Press, New York.

[16] Kumar, K.S. and Chandan, P. (2013). Risk-sensitive control of jump process on denumerable state space with near monotone cost. *Appl. Math. Optim.* **68**, 311–331.

[17] Patek, S.(2001). On terminating Markov decision processes with a risk-averse objective function. *Automatica* **37**, 1379-1386.

[18] Piunovski, A. and Khametov, V. (1985). New effective solutions of optimality equations for the controlled Markov chains with continuous parameter (the unbounded price-function). *Problems Control Inform. Theory* **14**, 303–318.

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

[20] Piunovskiy, A. and Zhang, Y. (2014). Discounted continuous-time Markov decision processes with unbounded rates and randomized history-dependent policies: the dynamic programming approach. *4OR-Q J. Operat. Res.* **12**, 4975.

[21] Wei, Q. (2016). Continuous-time Markov decision processes with risk-sensitive finite-horizon cost criterion. *Math. Meth. Oper. Res.* **84**, 461–487.

[22] Wei, Q. and Chen, X. (2016). Continuous-time Markov decision processes under the risk-sensitive average cost criterion. *Oper. Res. Lett.* **44**, 457–462.

[23] Zhang, Y. (2017). Continuous-time Markov decision processes with exponential utility. *SIAM J. Control Optim.* **55**, 2636-2660.