Haifeng Huo; Xian Wen

Risk probability optimization problem for finite horizon continuous time Markov decision processes with loss rate

Kybernetika, Vol. 57 (2021), No. 2, 272–294

Persistent URL: http://dml.cz/dmlcz/149039

KYBERNETIKA — VOLUME 57 (2021), NUMBER 2, PAGES 272–294

RISK PROBABILITY OPTIMIZATION PROBLEM FOR FINITE HORIZON CONTINUOUS TIME MARKOV DECISION PROCESSES WITH LOSS RATE

Haifeng Huo and Xian Wen

This paper presents a study the risk probability optimality for finite horizon continuous-time Markov decision process with loss rate and unbounded transition rates. Under drift condition, which is slightly weaker than the regular condition, as detailed in existing literature on the risk probability optimality Semi-Markov decision processes, we prove that the value function is the unique solution of the corresponding optimality equation, and demonstrate the existence of a risk probability optimization policy using an iteration technique. Furthermore, we provide verification of the imposed condition with two examples of controlled birth-and-death system and risk control, and further demonstrate that a value iteration algorithm can be used to calculate the value function and develop an optimal policy.

Keywords: continuous-time Markov decision processes, loss rate, risk probability criterion, finite horizon, optimal policy, unbounded transition rate

Classification: 90C40, 60E20

1. INTRODUCTION

Risk probability problems have formed a class of important stochastic optimization problems, that can be used in risk analysis, queueing systems and finance [4, 6, 9, 11, 13, 24, 32]. In contrast to the classical expected optimality problem [6, 7, 8, 20, 27] that focuses on the expectations of the total reward/costs, the risk probability optimization problem aims at minimizing (or maximizing) the risk probability, which means that the total loss (or reward) during a given time range is no more than (or exceeds) a given initial loss (or reward) goal. The results of this process can then be used to measure the risk of a stochastic system (economic and financial systems). Inspired by this situation, risk probability criteria have garnered significant attention and have been widely studied by [1, 2, 6, 10, 13, 15, 26, 28, 29, 31] for Markov decision processes (for short MDPs).

Risk probability optimality problems for Markov decision processes are first divided into three groups that are based on the hold times of the system state: discrete-time Markov decision processes (DTMDPs) [2, 26, 28, 29, 30, 31], semi-Markov decision processes (SMDPs) [10, 11, 12, 13, 25], and continuous-time Markov decision processes

DOI: 10.14736/kyb-2021-2-0272

(CTMDPs) [14, 15, 16]. Then the second classification is grouped by the risk probability optimization problems with the reward case or the loss case. Most of the earlier studies focus on the reward case, that minimizes the risk probability $P^{\pi}(B_r \leq \lambda)$ over all the policies π, where B_r denotes the total reward during a given time horizon, λ denote the reward level. It is clear that $P^{\pi}(B_r \leq \lambda) = 1 - P^{\pi}(B_r > \lambda)$, which combined with the conclusions from [13, 23] suggests that the risk probability minimization problem $P^{\pi}(B_r \leq \lambda)$ in [16] is not equivalent to the minimization problem $P^{\pi}(B_r > \lambda)$ in this paper. Moreover, in some control models such as economic and financial systems, the controller is often focused on the probability that the total loss incurred over a given time horizon exceeds the initial capacity. Hence, limited literature [13, 19] is available for the loss case, which minimizes the risk probability $P^{\pi}(B_l > \lambda)$ over all the policies π, where B_l denotes the total loss during a given time horizon, λ denotes the loss level (or goal). Specifically, Huang, Zou, and Guo [13] investigate the loss rates risk probability for first passage SMDPs, They use the invariant embedding technique to establish the optimality equation and prove the existence of optimal risk probability policies. Similarly, Liu and Zou [19] consider the risk probability criterion for finite horizon SMDPs with loss rate using the idea and iteration technique in [14] to demonstrate that the value function satisfies the optimality equation and the existence of optimal policies, and to derive an efficient algorithm for solving the value function. A review of the above mentioned literature demonstrates that the risk probability criterion with loss rate just considered in SMDPs, and CTMDPs for the risk probability criterion with the loss rate have not yet been explored. Moreover, there are many real-word situations, such as queueing systems, in which the lifetime is usually finite. Therefore, to the best of our knowledge, it is prudent to research the risk probability criterion for finite CTMDPs with loss rates in this paper.

Compared with the first passage risk probability CTMDPs developed in [15], the considered ones for finite-horizon CTMDPs with loss rates have many different characteristics due to their different performance criteria. (i) To define the policies, both loss levels λ and planning horizons t should be considered the extended states' components, while only reward levels λ have been considered in [15]. (ii) Our condition here is weaker than those proposed in [15]. The existence of optimal policies here is guaranteed by using the non-explosion of the controlled state process (see Assumption 1 in our paper), while the existence of optimal policies is guaranteed by using the non-explosion of the controlled state process and the properties of the target set B (see Assumption 3.2 and 3.6 in [15]). (iii) According to different policies, the probability space and the optimality equation in our paper are different from those developed in [15].

Since a key feature of our proposed model is that the loss levels and planning horizons are considered when the controller makes decisions. Thus, we first characterize the history-dependent policy with the system's states, loss levels and planning horizons, and reestablish a probability space. Secondly, following the same method utilized in [6, 7, 8], we establish a so-called drift condition to ensure the controlled state process is non-explosive. As a result of this condition and the continuity and compactness condition, we use an iteration technique to prove that the value function is the unique solution to the corresponding optimality equation, and from this optimization equation we demonstrate that a risk probability optimization policy exists. It should be noted

that the drift condition is imposed on the transaction rates for CTMDPs, which is typical, unlike the regular condition as explained in [10, 11, 12, 13, 25], because the transition rates are allowed to be unbounded, as detailed in Remark 3.3. Moreover, a value iteration algorithm is developed for calculating the value function and optimal policies. Finally, we provide two examples to explain our primary results. In the first example we demonstrate that our condition is verifiable via a controlled birth-and-death system. In the second example we use the value iterative algorithm to compute the value function and optimal policies.

The remainder of this paper is organized as follows. In Section 2, we describe the model of CTMDPs with the risk probability criterion, in which the policies depend on the states of the system, loss levels, and planning horizons. In Section 3, we present the value iteration technique for solving the concerned optimization problem. Our results are illustrated with two examples in Section 4.

2. OPTIMAL CONTROL PROBLEM

Continuous-time Markov decision processes (CTMDPs) model consists of the data

$$\{E, A, (A(i) \subseteq A, i \in E), q(j|i,a), c(i,a)\} \tag{1}$$

with the following interpretation:

(a) E denotes the state space, which is assumed to be a nonempty denumerable set;

(b) A denotes the action space, which is assumed to be a nonempty Borel space, endowed with the Borel σ-algebra $\mathcal{B}(A)$;

(c) $A(i)$ denotes a set of admissible actions at a given state $i \in E$. The subset $K := \{(i,a)|i \in E, a \in A(i)\}$ of $E \times A$ represents the set of allowed state-action pairs;

(d) $q(j|i,a)$ denote the transition rates, which are assumed to be conservative in that

$$\sum_{j \in E} q(j|i,a) = 0 \quad \forall (i,a) \in K, \tag{2}$$

and stable in the sense

$$q^*(i) = \sup_{a \in A(i)} q_i(a) < \infty \quad \forall (i,a) \in K, \tag{3}$$

where $q_i(a) := -q(i|i,a) \geq 0$ for all $(i,a) \in K$ and $q(j|i,a) \geq 0$ for all $(i,a) \in K$ such that $j \neq i$;

(e) $c(i,a)$ denotes the loss rate, which is assumed to be a nonnegative real-valued function on K.

The stochastic evolution of the control model (1) is described as follows: At the initial time $s_0 = 0$, the system is in the state i_0 and the controller has a loss level $\tilde{\lambda}_0 := \lambda_0$. The controller will do his/her best to manage the loss level during the planning horizon t_0. Roughly speaking, given the system state i_0, the loss level λ_0 and the planning horizon

t_0, the controller chooses an action $a_0 \in A(i_0)$ according to some given policy. Once such an action is taken, two things happen:

(i) The system stays at state i_0 until time s_1. At time s_1, the system moves to a new state i_1 with the probability $\frac{q(i_1|i_0,a_0)}{q_{i_0}(a_0)}(q_{i_0}(a_0) \neq 0)$. The holding time $\theta_1 = s_1 - s_0$ satisfies the exponential distribution $1 - e^{-q_{i_0}(a_0)(s_1-s_0)}$.

(ii) At time s_1, a loss $c(i_0, a_0)(s_1 - s_0)$ is incurred. Based on the current state i_1, loss level $\tilde{\lambda}_1 = [\lambda_0 - c(i_0, a_0)(s_1 - s_0)]^+$, planning horizon $t_1 = [t_0 - (s_1 - s_0)]^+$, and the previous state i_0, loss level λ_0, planning horizon t_0, the controller chooses another action $a_1 \in A(i_1)$, and the process is repeated. Thus, a sequence of losses stemming from all the taken actions will be incurred. The aim is to seek out a control policy that optimizes the risk probability criterion, that is the probability that the total losses exceeds a loss level during a fixed planning horizon.

To formalize the above description, we denote by s_k $(k \geq 1)$ the kth decision epoch, by i_k the state of the system on $[s_k, s_{k+1})$, by a_k the action at time s_k, by $\theta_k := s_k - s_{k-1}$ the holding time at state i_{k-1}. Moreover, we denote by $\tilde{\lambda}_k$ the loss level at the decision epoch s_k, and by t_k the planning horizon at the decision epoch s_k. The loss level $\tilde{\lambda}_k$ satisfies

$$\tilde{\lambda}_k := L_1(i_{k-1}, \tilde{\lambda}_{k-1}, a_{k-1}, \theta_k) := [\tilde{\lambda}_{k-1} - c(i_{k-1}, a_{k-1})\theta_k]^+, \tag{4}$$

and the planning horizon t_k satisfies

$$t_k := L_2(t_{k-1}, \theta_k) := [t_{k-1} - \theta_k]^+. \tag{5}$$

Here and everywhere else, the state process after moment $s_\infty := \lim_{k\to\infty} s_k$ is considered to remain in the artificial state $i_\infty := \Delta \notin E$ forever. Thus, we set $q(\cdot|\Delta, a_\Delta) := 0$, $c(\Delta, a_\Delta) := 0$, $A_\Delta := A \cup \{a_\Delta\}$ with isolated point a_Δ.

When the decision maker chooses actions, he/she must consider not only the state of the system, but also the loss level and the planning horizon.Thus, we have to redefine some policies and reestablish a probability space. The measurable space $(\Omega, \mathcal{F})$ is defined by

$$\begin{aligned}\Omega \;:=\; & \Omega^0 \bigcup \{(i_0, \lambda_0, t_0, s_1, i_1, \lambda_1, t_1, \ldots, s_k, i_k, \lambda_k, t_k, \ldots, \infty, \Delta, \infty, \infty, \ldots) | i_0 \in E, \lambda_0 \\ & \in [0, +\infty), t_0 \in [0, +\infty), s_l \in (0, \infty], i_l \in E, \lambda_l \in [0, +\infty), t_l \in [0, \infty), \forall 1 \leq l \leq \\ & k, k \geq 1\},\end{aligned}$$

and $\mathcal{F}$ denotes the Borel σ-algebra on Ω, where $\Omega^0 := E \times [0, +\infty) \times [0, +\infty) \times ((0, +\infty] \times E \times [0, +\infty) \times [0, +\infty))^\infty$. For each $k \geq 0$, $e := (i_0, \lambda_0, t_0, s_1, i_1, \lambda_1, t_1, \ldots, s_k, i_k, \lambda_k, t_k, \ldots) \in \Omega$, the k-component internal history is given by $h_0(e) := (i_0, \lambda_0, t_0)$, $h_k(e) := (i_0, \lambda_0, t_0, s_1, i_1, \lambda_1, t_1, \ldots, s_k, i_k, \lambda_k, t_k)$, the projections S_k, X_k, Λ_k, T_k are defined by

$$S_k(e) := s_k, \quad X_k(e) := i_k, \quad \Lambda_k(e) := \lambda_k, \quad T_k(e) := t_k,$$

and $S_\infty := \lim_{k\to\infty} S_k$. For simplicity, the argument e will often be omitted. The state process $\{x_s\}$ is defined by

$$x_s := \sum_{k\geq 0} I_{\{S_k \leq s < S_{k+1}\}} i_k + \Delta I_{\{s \geq S_\infty\}} \quad \text{for } s \geq 0, \tag{6}$$

where I_B denotes the indicator function of a set B.

Definition 2.1. A *deterministic history-dependent policy* is a sequence $\pi = \{f_0, f_1, \ldots\}$ of Borel measurable functions from Ω onto A_Δ for each $k = 0, 1, 2, \ldots$, and such that for each $e \in \Omega, s \geq 0$,

$$\pi(e, s) = I_{\{s=0\}} f_0(h_0(e)) + \sum_{k\geq 0} I_{\{S_k < s \leq S_{k+1}\}} f_k(h_k(e)) + I_{\{s \geq S_\infty\}} \delta_{a_\Delta}(da), \tag{7}$$

where $\delta_{a_\Delta}(da)$ denotes the Dirac measure on A_Δ concentrated on the isolated point a_Δ. The set of all deterministic history-dependent policies is denoted by Π.

A deterministic history-dependent policy $\pi = \{f_0, f_1, \ldots\}$ is said to be Markovian, if there exist some Borel measurable functions $\tilde{f}_k (k \geq 0)$ from $E \times R^+ \times [0, T]$ to A_Δ, such that $f_k(h_k(e)) = \tilde{f}_k(i_k, \lambda_k, t_k)$ for all $e \in \Omega$. We denote by Π_m the set of all deterministic Markov policies.

A deterministic Markov policy $\pi = \{\tilde{f}_0, \tilde{f}_1, \ldots\} \in \Pi_m$ is said to be stationary if there exists a Borel measurable function f from $E \times R^+ \times [0, T]$ to $A_\Delta$such that $\tilde{f}_k = f$. When it is the case, we denote by f such policy. The set of all deterministic stationary policies is denoted by Π_s. Clearly, we have $\Pi_s \subseteq \Pi_m \subseteq \Pi$.

For any given policy $\pi = \{f_0, f_1, \ldots\} \in \Pi$, from [7, 17], the jumps intensity of the process $\{x_s\}$ is given by

$$m^\pi(j|e, s) \quad = \quad I_{\{s=0\}} m_0^\pi(j|h_0(e)) + \sum_{k\geq 0} I_{\{S_k < s \leq S_{k+1}\}} m_k^\pi(j|h_k(e)), \tag{8}$$

where $m_0^\pi(j|h_0(e)) := q(j|i_0, f_0(h_0(e)))I_{\{j\neq i_0\}}$,$m_k^\pi(j|h_k(e)) := q(j|i_k, f_k(h_k(e)))I_{\{j\neq i_k\}}$.

Due to the changes in the loss levels and the planning horizons here, for each policy $\pi = \{f_0, f_1, \ldots\} \in \Pi$ and each initial probability measure γ on $E \times R^+ \times R^+$, according to the Ionescu Tulcea theorem (e. g., Proposition 7.45 in [3]), there exists a unique probability measure P_γ^π space on the measure space $(\Omega, \mathcal{F}, P_\gamma^\pi)$, which has a projection onto k-component internal history H_k with

$$P_{\gamma,0}^\pi(i, d\lambda_0, dt_0) \quad := \quad \gamma(i, d\lambda_0, dt_0), \tag{9}$$

$$\begin{aligned} P_{\gamma,k+1}^\pi(\Gamma \times (ds, j, d\lambda_{k+1}, dt_{k+1})) \quad := \quad & \int_\Gamma P_{\gamma,k}^\pi(dh_k) I_{\{\theta_k < \infty\}} m_k^\pi(j|h_k) \\ & \times \exp\{-m_k^\pi(E|h_k)(s - S_k)\} \delta_{L_2(t_k, s-S_k)}(dt_{k+1}) \\ & \times \delta_{L_1(i_k, \lambda_k, f_k(h_k), s-S_k)}(d\lambda_{k+1})\, \mathrm{d}s, \end{aligned} \tag{10}$$

$$\begin{aligned} P_\gamma^\pi(\Gamma \times (\infty, \Delta, \infty, \infty)) \quad := \quad & \int_\gamma P_\gamma^\pi(dh_k) I_{\{\theta_k = \infty\}} \\ & + I_{\{\theta_k < \infty\}} \exp\{-\int_0^\infty m_k^\pi(E|h_k)\, \mathrm{d}v\}, \end{aligned} \tag{11}$$

for $(i, d\lambda_0, dt_0) \in E \times \mathcal{B}(R^+) \times \mathcal{B}(R^+)$, where $H_0 := E \times R^+ \times R^+$ and $H_k := (E \times R^+ \times R^+) \times ((0, \infty] \times E_\Delta \times R^+ \times R^+)^k$, for $k = 1, 2, \ldots$, $\Gamma \in \mathcal{B}(H_k), m_k^\pi(E|h_k) := -q(i_k|i_k, f_k(h_k))$, and $\mathcal{B}(X)$ stands for the σ-algebra on X.

Let $\mathbb{E}^{\pi}_{\gamma}$ denotes the expectation operator with respect to P^{π}_{γ}. If the initial probability measure γ is concentrated on the initial $(i,\lambda,t) \in E \times R^{+} \times R^{+}$, we shall use $\mathbb{E}^{\pi}_{(i,\lambda,t)}$ and $P^{\pi}_{(i,\lambda,t)}$ instead of $\mathbb{E}^{\pi}_{\gamma}$ and P^{π}_{γ}, respectively.

To ensure the existence of the risk probability optimal policy, we need to assume that the state process $\{x_s, s \geq 0\}$ is nonexplosive.

Assumption 2.1. For any $\pi \in \Pi, (i,\lambda,t) \in E \times R^{+} \times [0,T]$, $P^{\pi}_{(i,\lambda,t)}(S_{\infty} = \infty) = 1$.

The main goal of Assumption 2.1 is to avoid the possibility of an infinite number of jumps during any finite horizon. We give a sufficient condition used in [5, 6, 7, 14] for the verification of Assumption 2.1.

Lemma 2.2. If there exist some constants $c_0 > 0$, $b_0 \geq 0$, $M_0 \geq 0$ and a measurable function $V \geq 1$ on E satisfying the following condition:

(a) $\sum_{j \in E} V(j) q(j \mid i,a) \leq c_0 V(i) + b_0$, for all $(i,a) \in K$;

(b) $q^{*}(i) \leq M_0 V(i)$ for all $i \in E$, with $q^{*}(i) = \sup_{a \in A(i)} q_i(a)$.

then Assumption 2.1 holds.

P r o o f . The statement follows from Theorem 1 in [14]. □

Remark 2.3. (1) Lemma 2.2 is a generalization of the drift condition introduced in [6, 7, 8]. The conditions of the Lemma 2.2 are satisfied when the transition rates are uniformly bounded (i. e. $\sup_{i \in E} q^{*}(i) < \infty$).

(2) It should be noted that there is a significant difference between Lemma 2.2 and the regular condition described in some existing literature on SMDPs [11, 12, 13]. The regular condition means that the semi-Markov kernel $Q(\delta, E|i,a)$ satisfies $Q(\delta, E|i,a) \leq 1 - \varepsilon$ for some constants $\delta > 0$ and $\varepsilon > 0$ and $(i,a) \in K$. For the model of CTMDPs, it means that $1 - e^{-q_i(a)\delta} \leq 1 - \varepsilon$, or equivalently, $e^{-q_i(a)\delta} \geq \varepsilon$ for all $(i,a) \in K$. This implies that the transition rates $q(j|i,a)$ must be bounded. Note that, in this work, we allow for the transition rates to be unbounded, see Example 4.1.

For any $(i,\lambda) \in E \times R^{+}$ and $\pi \in \Pi$, the risk probability criterion $U^{\pi}(i,\lambda,T)$ with loss rate on the finite horizon T is given by

$$U^{\pi}(i,\lambda,T) := P^{\pi}_{(i,\lambda,T)}\Big(\int_0^T c(x_s, \pi_s)\, \mathrm{d}s > \lambda \Big), \tag{12}$$

where $c(x_s,\pi_s)(e) := c(x_s(e), \pi(e,s))$ for all $e \in \Omega$ and $s \geq 0$.

Definition 2.4. A policy $\pi^{*} \in \Pi$ is called risk probability optimal if

$$U^{\pi^{*}}(i,\lambda,T) = U^{*}(i,\lambda,T) \quad \forall (i,\lambda) \in E \times R^{+}, \tag{13}$$

where $U^{*}(i,\lambda,T) := \inf_{\pi \in \Pi} U^{\pi}(i,\lambda,T)$ is the corresponding risk probability value function.

3. MAIN RESULTS

In this section, some suitable conditions are provided for ensuring the existence of optimality equation and the optimal policies. Moreover, an efficient method is developed for computing the value function and the optimal policies.

Notation: For a policy $\pi \in \Pi$, let $U^\pi(i,\lambda,t)$ be the corresponding risk probability of the controlled system from 0 to time $t \in [0,T]$, given the initial state $i \in E$ and the loss level $\lambda \in R^+$, i. e.,

$$U^\pi(i,\lambda,t) := P^\pi_{(i,\lambda,t)}\Big(\int_0^t c(x_s,\pi_s)\,\mathrm{d}s > \lambda\Big). \tag{14}$$

Let

$$U^*(i,\lambda,t) = \inf_{\pi\in\Pi} U^\pi(i,\lambda,t) \quad \forall (i,\lambda,t) \in E \times R^+ \times [0,T]. \tag{15}$$

Denote by $\mathcal{U}_m$ the set of all Borel measurable functions from $E \times R^+ \times [0,T]$ to $[0,1]$.

For each $(i,\lambda,t) \in E \times R^+ \times [0,T], U \in \mathcal{U}_m, f \in \Pi_s$ and $a \in A(i)$, we define the operators $H^f U$ and HU on $\mathcal{U}_m$ by

$$\begin{aligned} H^f U(i,\lambda,t) &:= I_{(\lambda,+\infty)}(c(i,f)t)e^{-q_i(f)t} \\ &\quad + \sum_{j\neq i}\int_0^t U\Big(j,\lambda - c(i,f)u, t-u\Big)e^{-q_i(f)u}q(j|i,f)\,\mathrm{d}u \end{aligned} \tag{16}$$

$$\begin{aligned} H^a U(i,\lambda,t) &:= I_{(\lambda,+\infty)}(c(i,a)t)e^{-q_i(a)t} \\ &\quad + \sum_{j\neq i}\int_0^t U\Big(j,\lambda - c(i,a)u, t-u\Big)e^{-q_i(a)u}q(j|i,a)\,\mathrm{d}u \end{aligned} \tag{17}$$

$$HU(i,\lambda,t) := \inf_{a\in A(i)} H^a U(i,\lambda,t), \tag{18}$$

with $q_i(f) := -q(i|i,f(i,\lambda,t))$ and $q(j|i,f) := q(j|i,f(i,\lambda,t))$.

Similarly, for every $f \in \Pi_s$, we define iteratively the operators $(H^n U, n \geq 1), ((H^f)^n U, n \geq 1)$ on $\mathcal{U}_m$ by setting

$$H^1 U = HU, H^{n+1}U = H(H^n U), (H^f)^1 U = H^f U, (H^f)^{n+1}U = H^f((H^f)^n U), n \geq 1.$$

From the theoretical perspective, to show the existence of a risk probability optimal policy, as in CTMDPs [6, 7, 8, 11], we introduce the following assumption.

Assumption 3.1. For any fixed $(i,\lambda,t) \in E \times R^+ \times [0,T]$,

(a) $A(i)$ is compact.

(b) For all $i,j \in E$, the function $c(i,a)$ and $q(j|i,a)$ are continuous in $a \in A(i)$ and $q(i|i,a)$ is inf-compact on K.

(c) For each fixed $U \in \mathcal{U}_m$, $\sum_{j\neq i}\int_0^t U\Big(j, \lambda - c(i,a)u, t-u\Big)e^{-q_i(a)u}q(j|i,a)\,\mathrm{d}u$ is lower semicontinuous in $a \in A(i)$.

Remark 3.1. Assumption 3.1 is so-called continuity-compactness condition, which is trivially satisfied when the action space is denumerable and the set $A(i)$ is finite for all $i \in E$.

We list below some important properties of these operators.

Lemma 3.2. Suppose that Assumption 2.1 and 3.1 hold. The following assertions hold:

(a) If $U, V \in \mathcal{U}_m$ with $U(i,\lambda,t) \geq V(i,\lambda,t)$ for all $(i,\lambda,t) \in E \times R^+ \times [0,T]$, then $H^aU(i,\lambda,t) \geq H^aV(i,\lambda,t)$ and $HU(i,\lambda,t) \geq HV(i,\lambda,t)$, for every $a \in A(i)$.

(b) If $U \in \mathcal{U}_m$, then $HU \in \mathcal{U}_m$, and there exists a policy $f \in \Pi_s$ for which the infimum in (18) is attained at $f(i,\lambda,t) \in A(i)$, i. e.,

$$HU(i,\lambda,t) = H^fU(i,\lambda,t) \quad \forall (i,\lambda,t) \in E \times R^+ \times [0,T]. \tag{19}$$

P r o o f . (a) This part follows from the definition of the operator H.

(b) If $U \in \mathcal{U}_m$, by (18), we know that $HU \in \mathcal{U}_m$. Moreover, under Assumption 2.1 and 3.1, for any $(i,\lambda,t) \in E \times R^+ \times [0,T]$, by the measurable selection theorem (Proposition D.5(a) in [9]), we know that there is a policy $f \in \Pi_s$ for which the infimum in (18) is attained at $f(i,\lambda,t) \in A(i)$. □

For any $(i,\lambda,t) \in E \times R^+ \times [0,T]$ and $\pi \in \Pi$, and based on facts that the state process $\{x_s, s \geq 0\}$ is non-explosive, the loss rate $c(i,a)$ is nonnegative, and the probability measure is continuous, we may write $U^\pi(i,\lambda,t)$ as

$$\begin{aligned}
U^\pi(i,\lambda,t) &= P^\pi_{(i,\lambda,t)}\Big(\int_0^t c(x_s,\pi_s)\,\mathrm{d}s > \lambda\Big) \\
&= P^\pi_{(i,\lambda,t)}\Big(\sum_{m=0}^{\infty}\int_{S_m\wedge t}^{S_{m+1}\wedge t} c(x_s,\pi_s)\,\mathrm{d}s > \lambda\Big) \\
&= P^\pi_{(i,\lambda,t)}\Big(\bigcap_{n=1}^{\infty}\sum_{m=0}^{n}\int_{S_m\wedge t}^{S_{m+1}\wedge t} c(x_s,\pi_s)\,\mathrm{d}s > \lambda\Big) \\
&= \lim_{n\to\infty} P^\pi_{(i,\lambda,t)}\Big(\sum_{m=0}^{n}\int_{S_m\wedge t}^{S_{m+1}\wedge t} c(x_s,\pi_s)\,\mathrm{d}s > \lambda\Big) \\
&:= \lim_{n\to\infty} U^\pi_n(i,\lambda,t).
\end{aligned}$$

Thus, we obtain a sequence $\{U^\pi_n(i,\lambda,t), n = -1, 0, 1, \ldots\}$ with $U^\pi_{-1}(i,\lambda,t) := 0$, satisfying $0 \leq U^\pi_n(i,\lambda,t) \leq U^\pi_{n+1}(i,\lambda,t) \leq 1, n \geq -1$ and $\lim_{n\to\infty} U^\pi_n(i,\lambda,t) = U^\pi(i,\lambda,t)$.

The following lemma provides a fundamental result for solving the optimality equation.

Lemma 3.3. Suppose that Assumption 2.1 and 3.1 hold. Then, for each $(i,\lambda,t) \in E \times R^+ \times [0,T]$, $\pi = \{f_0, f_1, \ldots\} \in \Pi$ and $n \geq -1$,

(a) $U_n^\pi \in \mathcal{U}_m$ and $U^\pi \in \mathcal{U}_m$.

(b) $U_{n+1}^\pi(i,\lambda,t) = H^{f_0}U_n^{^1\pi}(i,\lambda,t)$, and $U^\pi(i,\lambda,t) = H^{f_0}U^{^1\pi}(i,\lambda,t)$ with the 1-shift policy of π, i. e. $\pi^1 := (\hat{f}_0, \hat{f}_1, ...)$, $\hat{f}_k(s_1, i_1, \lambda_1, t_1, \ldots, s_{k+1}, i_{k+1}, \lambda_{k+1}, t_{k+1}) := f_{k+1}(i,\lambda,t,s_1,i_1,\lambda_1,t_1,\ldots,s_{k+1},i_{k+1},\lambda_{k+1},t_{k+1}), k = 0,1,\ldots$.

In particular, for $f \in \Pi_s$, $U_{n+1}^f(i,\lambda,t) = H^f U_n^f(i,\lambda,t)$ and $U^f(i,\lambda,t) = H^f U^f(i,\lambda,t)$.

P r o o f. (a) We shall prove part(a) by induction on the integer n. The claims being obvious for $n = -1$ since $U_{-1}^\pi(i,\lambda,t) = 0 \in \mathcal{U}_m$ for $(i,\lambda,t) \in E \times R^+ \times [0,T], \pi \in \Pi$. We assume they hold for any $n \geq -1$. From (10) and the property of the conditional expectation, we have

$$
\begin{aligned}
U_{n+1}^\pi(i,\lambda,t) &= P_{(i,\lambda,t)}^\pi\Big(\sum_{m=0}^{n+1}\int_{S_m\wedge t}^{S_{m+1}\wedge t} c(x_s,\pi_s)\,\mathrm{d}s > \lambda\Big)\\
&= P_{(i,\lambda,t)}^\pi\Big(\int_0^t c(x_s,\pi_s)\,\mathrm{d}s > \lambda, S_1 > t\Big)\\
&\quad + P_{(i,\lambda,t)}^\pi\Big(\sum_{m=0}^{n+1}\int_{S_m\wedge t}^{S_{m+1}\wedge t} c(x_s,\pi_s)\,\mathrm{d}s > \lambda, S_1 \leq t\Big)\\
&= E_{(i,\lambda,t)}^\pi[I_{\{\int_0^t c(x_s,\pi_s)\,\mathrm{d}s>\lambda, S_1>t\}}]\\
&\quad + E_{(i,\lambda,t)}^\pi[I_{\{\sum_{m=0}^{n+1}\int_{S_m\wedge t}^{S_{m+1}\wedge t} c(x_s,\pi_s)\,\mathrm{d}s>\lambda, S_1\leq t\}}]\\
&= E_{(i,\lambda,t)}^\pi[E_{(i,\lambda,t)}^\pi[I_{\{\int_0^t c(x_s,\pi_s)\,\mathrm{d}s>\lambda, S_1>t\}}|S_1, x_{S_1}, \Lambda_1, T_1]]\\
&\quad + E_{(i,\lambda,t)}^\pi[E_{(i,\lambda,t)}^\pi[I_{\{\sum_{m=0}^{n+1}\int_{S_m\wedge t}^{S_{m+1}\wedge t} c(x_s,\pi_s)\,\mathrm{d}s>\lambda, S_1\leq t\}}|S_1, x_{S_1}, \Lambda_1, T_1]]\\
&= \sum_{j\neq i}\int_0^{+\infty} P_{(i,\lambda,t)}^\pi\Big(\int_0^t c(x_s,\pi_s)\,\mathrm{d}s > \lambda, S_1 > t | S_1 = u, x_{S_1} = j,\\
&\quad \Lambda_1 = \lambda - c(i,f_0)u, T_1 = [t-u]^+\Big) e^{-q_i(f_0)u} q(j|i,f_0)\,\mathrm{d}u\\
&\quad + \sum_{j\neq i}\int_0^{+\infty} P_{(i,\lambda,t)}^\pi\Big(\int_0^u c(x_s,\pi_s)\,\mathrm{d}s + \sum_{m=1}^{n+1}\int_{S_m\wedge t}^{S_{m+1}\wedge t} c(x_s,\pi_s)\,\mathrm{d}s\\
&\quad > \lambda, S_1 \leq t | S_1 = u, x_{S_1} = j, \Lambda_1 = \lambda - c(i,f_0)u, T_1 = [t-u]^+\Big)\\
&\quad \times e^{-q_i(f_0)u} q(j|i,f_0)\,\mathrm{d}u\\
&= I_{(\lambda,\infty)}(c(i,f_0)t)e^{-q_i(f_0)t} + \sum_{j\neq i}\int_0^t P_{(i,\lambda,t)}^\pi\Big(\sum_{m=1}^{n+1}\int_{S_m\wedge t}^{S_{m+1}\wedge t} c(x_s,\pi_s)\,\mathrm{d}s\\
&\quad > \lambda - c(i,f_0)u, S_1 \leq t | S_1 = u, x_{S_1} = j,\\
&\quad \Lambda_1 = \lambda - c(i,f_0)u, T_1 = [t-u]^+\Big) e^{-q_i(f_0)u} q(j|i,f_0)\,\mathrm{d}u\\
&= I_{(\lambda,+\infty)}(c(i,f_0)t)e^{-q_i(f_0)t}
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{j\neq i}\int_0^t P^{\pi}_{(i,\lambda,t)}\Big(\sum_{m=1}^{n+1}\int_{(S_m\wedge t)-u}^{(S_{m+1}\wedge t)-u} c(x_{l+u},\pi_{l+u})\,\mathrm{d}l > \lambda - c(i,f_0)u,\\
&S_1\le t|S_1=u,x_{S_1}=j,\Lambda_1=\lambda-c(i,f_0)u,T_1=[t-u]^+\Big)\\
&\times e^{-q_i(f_0)u}q(j|i,f_0)\,\mathrm{d}u\\
=\;& I_{(\lambda,+\infty)}(c(i,f_0)t)e^{-q_i(f_0)t}\\
&+\sum_{j\neq i}\int_0^t P^{{}^1\pi}_{(j,\lambda-c(i,f_0)u,t-u)}\Big(\sum_{m=0}^{k}\int_{S_m\wedge(t-u)}^{S_{m+1}\wedge(t-u)} c(x_s,\pi_s)\,\mathrm{d}s\\
&>\lambda-c(i,f_0)u\Big)e^{-q_i(f_0)u}q(j|i,f_0)\,\mathrm{d}u\\
=\;& I_{(\lambda,+\infty)}(c(i,f_0)t)e^{-q_i(f_0)t}+\sum_{j\neq i}\int_0^t U_k^{{}^1\pi}\Big(j,\lambda-c(i,f_0)u,t-u\Big)\\
&\times e^{-q_i(f_0)u}q(j|i,f_0)\,\mathrm{d}u\\
:=\;& H^{f_0}U_k^{{}^1\pi}(i,\lambda,t).
\end{aligned}
$$

The induction hypothesis then gives $U^{\pi}_{n+1} := H^{f_0}U_k^{{}^1\pi} \in \mathcal{U}_m$. Thus, by the limit of a sequence of measurable functions is still measurable, we have $\lim_{n\to\infty} U^{\pi}_n = U^{\pi} \in \mathcal{U}_m$.

(b) Given $(i,\lambda,t) \in E\times R^+\times[0,T]$, we have $U^{\pi}_{n+1}(i,\lambda,t) = H^{f_0}U_n^{{}^1\pi}(i,\lambda,t)$, for any $n \ge -1$. Thus letting $n\to\infty$, and applying the dominated convergence theorem, we have $U^{\pi}(i,\lambda,t) = H^{f_0}U^{{}^1\pi}(i,\lambda,t)$. In particular, if $\pi = f \in \Pi_s$, then $U^f(i,\lambda,t) = H^fU^f(i,\lambda,t)$. □

Remark 3.4. Lemma 3.3 provides an efficient method for computing the function $U^f(i,\lambda,t)$, namely we have $U^f_{n+1}(i,\lambda,t) = H^fU^f_n(i,\lambda,t)$ *and* $U^f(i,\lambda,t) = \lim_{n\to\infty} U^f_n(i,\lambda,t)$ for any $(i,\lambda,t)\in E\times R^+\times[0,T], f\in\Pi_s$, where $U^f_{-1}(i,\lambda,t) := 0$.

The following result is new, and we shall use it to prove the uniqueness of the solution to the corresponding optimality equation.

Theorem 3.5. Suppose that Assumption 2.1 and 3.1 are satisfied.

(a) Given $U,V\in\mathcal{U}_m$ *and* $f\in\Pi_s$, if $U(i,\lambda,t)-V(i,\lambda,t)\le H^f(U-V)(i,\lambda,t)$, then $U(i,\lambda,t)\le V(i,\lambda,t)$ for all $(i,\lambda,t)\in E\times R^+\times[0,T]$.

(b) U^f is the unique solution in $\mathcal{U}_m$ to the equation

$$U(i,\lambda,t) = H^fU(i,\lambda,t)\quad \forall f\in\Pi_s,(i,\lambda,t)\in E\times R^+\times[0,T].$$

Proof. (*a*) To prove part (*a*), we shall show that for all $f\in\Pi_s,(i,\lambda,t)\in E\times R^+\times[0,T], n=1,2,\ldots,$

$$(H^f)^n(U-V)(i,\lambda,t)\le P^f_{(i,\lambda,t)}(S_n\le t). \tag{20}$$

We shall prove (3.7) by induction on the integer n. We know that

$$\begin{aligned}
& P^f_{(i,\lambda,t)}(S_1 \le t) \\
= \;& E^f_{(i,\lambda,t)}[I_{\{S_1 \le t\}}] \\
= \;& E^f_{(i,\lambda,t)}[E^f_{(i,\lambda,t)}[I_{\{S_1 \le t\}} | S_1, X_{S_1}, \Lambda_1, T_1]] \\
= \;& \sum_{j \ne i} \int_0^{+\infty} P^f_{(i,\lambda,t)}\Big(S_1 \le t | S_1 = u, X_{S_1} = j, \Lambda_1 = \lambda - c(i,f)u, \\
& T_1 = [t-u]^+\Big) e^{-q_i(f)u} q(j|i,f)\,\mathrm{d}u \\
= \;& \sum_{j \ne i} \int_0^{t} e^{-q_i(f)u} q(j|i,f)\,\mathrm{d}u \\
= \;& 1 - e^{-q_i(f)t}. \qquad (21)
\end{aligned}$$

Since $U, V \in \mathcal{U}_m$, by the definition of the operator H, we have

$$\begin{aligned}
& H^f(U-V)(i,\lambda,t) \\
= \;& \sum_{j \ne i} \int_0^t (U-V)\Big(j, \lambda - c(i,f)u, t-u\Big) \\
& \times e^{-q_i(f)u} q(j|i,f)\,\mathrm{d}u \\
\le \;& \sum_{j \ne i} \int_0^t e^{-q_i(f)u} q(j|i,f)\,\mathrm{d}u \\
= \;& P^f_{(i,\lambda,t)}(S_1 \le t),
\end{aligned}$$

where the last equality follows from (21). Thus, the statement (20) is true for $n = 1$. Suppose that the statement is true for any $n \ge 1$. By (10) and the property of the conditional expectation, we have

$$\begin{aligned}
& P^f_{(i,\lambda,t)}(S_{n+1} \le t) \\
= \;& E^f_{(i,\lambda,t)}[I_{\{S_{n+1} \le t\}}] \\
= \;& E^f_{(i,\lambda,t)}[E^f_{(i,\lambda,t)}[I_{\{S_{n+1} \le t\}} | S_1, X_{S_1}, \Lambda_1, T_1]] \\
= \;& \sum_{j \ne i} \int_0^{+\infty} P^f_{(i,\lambda,t)}\Big(S_{n+1} \le t | S_1 = u, X_{S_1} = j, \Lambda_1 = \lambda - c(i,f)u, \\
& T_1 = [t-u]^+\Big) e^{-q_i(f)u} q(j|i,f)\,\mathrm{d}u \\
= \;& \sum_{j \ne i} \int_0^{t} P^f_{(j,\lambda - c(i,f)u, t-u)}\Big(S_n \le t-u\Big) e^{-q_i(f)u} q(j|i,f)\,\mathrm{d}u. \qquad (22)
\end{aligned}$$

Moreover, from the definition of the operators $(H^f)^n$ and the induction hypothesis,

we obtain

$$\begin{aligned}
& (H^f)^{n+1}(U-V)(i,\lambda,t) \\
= \; & H^f(H^f)^n(U-V)(i,\lambda,t) \\
= \; & \sum_{j\neq i}\int_0^t (H^f)^n(U-V)\Big(j,\lambda-c(i,f)u,t-u\Big) \\
& \times e^{-q_i(f)u}q(j|i,f)\,\mathrm{d}u \\
\leq \; & \sum_{j\neq i}\int_0^t P^f_{(j,\lambda-r(i,f)u,t-u)}\Big(S_n\leq t-u\Big) \\
& \times e^{-q_i(f)u}q(j|i,f)\,\mathrm{d}u \\
= \; & P^f_{(i,\lambda,t)}(S_{n+1}\leq t)
\end{aligned}$$

where the last equality follows from (22). This together with the sufficient condition $U(i,\lambda,t)-V(i,\lambda,t)\leq H^f(U-V)(i,\lambda,t)$ and the induction hypothesis yield

$$U(i,\lambda,t)-V(i,\lambda,t)\leq (H^f)^n(U-V)(i,\lambda,t)\leq P^f_{(i,\lambda,t)}(S_n\leq t). \tag{23}$$

Letting $n\to\infty$ in (23), and taking into account Assumption 2.1, we obtain

$$U(i,\lambda,t)-V(i,\lambda,t)\leq \lim_{n\to\infty} P^f_{(i,\lambda,t)}(S_n\leq t)=0.$$

This concludes the proof of part (a) of the theorem.

(*b*) For each $(i,\lambda,t)\in E\times R^+\times[0,T]$, it follows from Lemma 3.2 (b) that $U^f(i,\lambda,t)=H^fU^f(i,\lambda,t)$. If V is another solution in $\mathcal{U}_m$ to the equation $U(i,\lambda,t)=H^fU(i,\lambda,t)$ for $(i,\lambda,t)\in E\times R^+\times[0,T]$, then $U^f(i,\lambda,t)-V(i,\lambda,t)=H^f(U^f(i,\lambda,t)-V(i,\lambda,t))$, which together with part (a) gives $U^f(i,\lambda,t)=V(i,\lambda,t)$. This concludes the proof. □

The following theorem is our main result. It shows that the unique solution to the corresponding optimality equation is the value function and proves the existence of optimal policies.

Theorem 3.6. Suppose that Assumption 2.1 and 3.1 hold.

(a) For any $(i,\lambda,t)\in E\times R^+\times[0,T]$ and $n\geq -1$, define

$$U^*_{n+1}:=HU^*_n,\quad \text{with}\quad U^*_{-1}:=0.$$

Then, $\lim_{n\to\infty}U^*_n=U^*$.

(b) The value function U^* is the unique solution in $\mathcal{U}_m$ to the optimality equation. i. e.,

$$U^*(i,\lambda,t)=HU^*(i,\lambda,t)\quad \forall (i,\lambda,t)\in E\times R^+\times[0,T]. \tag{24}$$

(c) There exists a policy $f^*\in\Pi_s$ with $U^*(i,\lambda,t)=H^{f^*}U^*(i,\lambda,t)$ and $U^*(i,\lambda,t)=U^{f^*}(i,\lambda,t)$.

(d) Let $\tilde{f}_0(i,\lambda,T) := f^*(i,\lambda,T)$, $\tilde{f}_k(i,\lambda,T,s_1,i_1,\lambda_1,t_1,\ldots,S_k,i_k,\lambda_k,t_k) := f^*(i_k,\tilde{\lambda}_k,t_k)$ for $(i,\lambda,T,\ s_1,i_1,\lambda_1,t_1,\ldots,S_k,i_k,\lambda_k,t_k) \in H_k, k \geq 1$. Then, The deterministic history-dependent policy $\pi^* := (\tilde{f}_0,\tilde{f}_1,\ldots,\tilde{f}_k)$ is optimal, where $\tilde{\lambda}_k = L_1(i_{k-1},\tilde{\lambda}_{k-1}, f^*(i_{k-1},\ \tilde{\lambda}_{k-1},\ t_{k-1}),\theta_k), t_k = L_2(t_{k-1},\theta_k), i_0 = i, \lambda_0 = \lambda, t_0 = T, \theta_k = s_k - s_{k-1}$, and L_1, L_2 are defined in (4),(5).

P r o o f. (a) By definition of the operator H and U_n^*, we have

$$0 \leq U_n^*(i,\lambda,t) \leq U_{n+1}^*(i,\lambda,t) \leq 1, n \geq -1.$$

This implies that $\lim_{n\to\infty} U_n^*(i,\lambda,t) := \tilde{U}(i,\lambda,t)$. To finish the proof, it remains to prove that $\tilde{U} = U^*$.

We first prove, for any $\pi \in \Pi$, by induction on the integer $n \geq -1$ that

$$U_n^*(i,\lambda,t) \leq U_n^\pi(i,\lambda,t) \ \ \forall (i,\lambda,t) \in E \times R^+ \times [0,T]. \tag{25}$$

Obviously, (25) holds for $n = -1$, since $U_{-1}^*(i,\lambda,t) = 0 = U_{-1}^\pi(i,\lambda,t)$ for any $\pi \in \Pi$. Suppose that $U_n^*(i,\lambda,t) \leq U_n^\pi(i,\lambda,t)$ for all $\pi = \{f_0,f_1,\ldots\} \in \Pi$ and $n \geq -1$. Then, by the induction hypothesis and Lemma 3.3(b), we have

$$U_{n+1}^*(i,\lambda,t) = HU_n^*(i,\lambda,t) \leq HU_n^{{}^1\pi}(i,\lambda,t) \leq H^{f^0}U_n^{{}^1\pi}(i,\lambda,t) = U_{n+1}^\pi(i,\lambda,t).$$

Letting $n \to \infty$, we obtain $\tilde{U}(i,\lambda,t) = \lim_{n\to\infty} U_n^*(i,\lambda,t) \leq U^\pi(i,\lambda,t)$. This implies that $\tilde{U}(i,\lambda,t) \leq U^*(i,\lambda,t)$, as π is arbitrary.

To prove the converse, we consider the sets $A_n := \{a \in A(i) | H^a U_n^*(i,\lambda,t) \leq H\tilde{U}(i,\lambda,t)\}$ for all $n \geq -1$ and $A^* := \{a \in A(i) | H^a\tilde{U}(i,\lambda,t) = H\tilde{U}(i,\lambda,t)\}$. Under Assumption 3.1, using the operator H is monotone, and $U_n^* \uparrow \tilde{U}$, we deduce that the sets A_n and A^* are nonempty and compact, and $A_n \downarrow A^*$. Hence, it follows from the measurable selection theorem (Proposition D.5(a) in [9]) that there exist $a_n \in A_n$ such that $H^{a_n}U_n^*(i,\lambda,t) = HU_n^*(i,\lambda,t)$. Thus, using the facts that $A(i)$ is compact and $A_n \downarrow A^*$, we conclude the existence of a subsequence $\{a_{n_k}\}$ of $\{a_n\}$ and $a^* \in A^*$ satisfying $a_{n_k} \to a^*$ as $n_k \to \infty$. Hence, in view of the monotonicity of the sequence $\{U_n^*\}$ and the definition of the operator H, we have

$$H^{a_{n_k}}U_{n_k}^*(i,\lambda,t) \geq H^{a_{n_k}}U_n^*(i,\lambda,t) \ \ \forall n_k \geq n.$$

Letting $k \to \infty$ and applying the dominated convergence theorem, we obtain

$$\lim_{k\to\infty} H^{a_{n_k}}U_{n_k}^*(i,\lambda,t) = \lim_{k\to\infty} HU_{n_k}^*(i,\lambda,t) \geq H^{a^*}U_n^*(i,\lambda,t).$$

Thus $\tilde{U}(i,\lambda,t) \geq H^{a^*}U_n^*(i,\lambda,t)$.

Furthermore, letting $n \to \infty$, we obtain that $\tilde{U}(i,\lambda,t) \geq H\tilde{U}(i,\lambda,t)$. The Lemma 3.2 (b) ensures the existence of a policy $f \in \Pi_s$ satisfying

$$\tilde{U}(i,\lambda,t) \geq H\tilde{U}(i,\lambda,t) = H^f\tilde{U}(i,\lambda,t),$$

which together with Lemma 3.3(b) and Remark 3.4 gives that

$$\tilde{U}(i,\lambda,t) \geq (H^f)^n\tilde{U}(i,\lambda,t) \geq (H^f)^n\tilde{U}_{-1}(i,\lambda,t) = U_{n-1}^f(i,\lambda,t).$$

Letting $n \to \infty$, we obtain that

$$\tilde{U}(i,\lambda,t) \geq U^{f}(i,\lambda,t) \geq U^{*}(i,\lambda,t).$$

This fact concludes the proof of part (a).

(b) By Lemma 3.3(b), for each $(i,\lambda,t) \in E \times R^{+} \times [0,T]$, we have

$$U^{\pi}(i,\lambda,t) = H^{f_0}U^{1\pi}(i,\lambda,t) \geq H^{f_0}U^{*}(i,\lambda,t) \geq HU^{*}(i,\lambda,t) \ \ \forall \pi \in \Pi,$$

This implies that $U^{*}(i,\lambda,t) \geq HU^{*}(i,\lambda,t)$, as π is arbitrary.

On the other hand, from part (a) and the definition of U_n^{*}, for each $(i,\lambda,t) \in E \times R^{+} \times [0,T]$, we have

$$U_{n+1}^{*}(i,\lambda,t) = HU_n^{*}(i,\lambda,t) \leq H^{a}U_n^{*}(i,\lambda,t) \ \ \forall a \in A(i).$$

Letting $n \to \infty$ and invoking the dominated convergence theorem, give

$$U^{*}(i,\lambda,t) \leq H^{a}U^{*}(i,\lambda,t).$$

Therefore, $U^{*}(i,\lambda,t) \leq HU^{*}(i,\lambda,t)$, since a was taken to be arbitrary. Thus, $U^{*} = HU^{*}$.

Furthermore, for each $(i,\lambda,t) \in E \times R^{+} \times [0,T]$, the existence of a policy $f \in \Pi_s$ satisfying $U^{*}(i,\lambda,t) = H^{f}U^{*}(i,\lambda,t)$ is ensured by Lemma 3.2(b). Similarly, if $V \in \mathcal{U}_m$ satisfies the equation $V(i,\lambda,t) = HV(i,\lambda,t)$ then there exists a policy $f^{'} \in \Pi_s$ such that $V(i,\lambda,t) = H^{f^{'}}V(i,\lambda,t)$. This implies that

$$U^{*}(i,\lambda,t) = H^{f}U^{*}(i,\lambda,t) \leq H^{f^{'}}U^{*}(i,\lambda,t), \tag{26}$$

$$V(i,\lambda,t) = H^{f^{'}}V(i,\lambda,t) \leq H^{f}V(i,\lambda,t). \tag{27}$$

Combining (26) and (27) gives

$$U^{*}(i,\lambda,t) - V(i,\lambda,t) \leq H^{f^{'}}(U^{*} - V)(i,\lambda,t), \tag{28}$$

$$V(i,\lambda,t) - U^{*}(i,\lambda,t) \leq H^{f}(U^{*} - V)(i,\lambda,t), \tag{29}$$

which together with Theorem 3.5(a) yield $U^{*}(i,\lambda,t) = V(i,\lambda,t)$. This proves part (b).

(c)According to Lemma 3.2(b), there exists a policy $f^{*} \in \Pi_s$ such that

$$U^{*}(i,\lambda,t) = H^{f^{*}}U^{*}(i,\lambda,t),$$

which together with Theorem 3.5(b) and part (b) gives $U^{*}(i,\lambda,t) = U^{f^{*}}(i,\lambda,t)$.

(d) Using (4),(7),(10), and the definition of π^{*} for any $(i,\lambda,t) \in E \times R^{+} \times [0,T]$, we have $P_{\gamma,k}^{f^{*}} = P_{\gamma,k}^{\pi^{*}}$ and for all $k \geq 0$. Therefore, $P_{\gamma}^{f^{*}} = P_{\gamma}^{\pi^{*}}$. Hence, we conclude that $U^{\pi^{*}}(i,\lambda,t) = U^{f^{*}}(i,\lambda,t) = U^{*}(i,\lambda,t)$, and so π^{*} is optimal. □

The arguments of Theorem 3.6 lead to the following iterative scheme for computing the value function $U^{*}(i,\lambda,t)$.

The Value iteration algorithm

Step 1: Set $U_{-1}^*(i,\lambda,t) := 0$ with $(i,\lambda,t) \in E \times R^+ \times [0,T]$.
Step 2: Substitute $n+1$ for n, set

$$\begin{aligned} H^a U_n^*(i,\lambda,t) &= I_{(\lambda,+\infty)}(c(i,a)t)e^{-q_i(a)t} \\ &\quad + \sum_{j\neq i} \int_0^t U_n^*\Big(j, \lambda - c(i,a)u, t-u\Big) e^{-q_i(a)u} q(j|i,a)\,\mathrm{d}u, \qquad (30) \\ U_{n+1}^*(i,\lambda,t) &= \min_{a\in A(i)} \{H^a U_n^*(i,\lambda,t)\}. \qquad (31) \end{aligned}$$

Step 3: For a given fully small positive ε, if $|U_{n+1}^*(i,\lambda,t) - U_n^*(i,\lambda,t)| < \varepsilon$, stop. Otherwise, return to step 2.

Since, for large value of n, $U_{n+1}^*(i,\lambda,t)$ and $U_n^*(i,\lambda,t)$ are highly close, the value U_{n+1}^* is usually taken as the value function U^*.

Remark 3.7. It should be noted that the formula (30) is derived from the trapezoidal integration method in [18], which is explained as below.

$$\int_a^b g(x)\,\mathrm{d}x \approx \sum_{k=0}^{m-1} \frac{g(a+kh) + g(a+(k+1)h)}{2} h, \qquad (32)$$

in which h is the step length, $k \leq m, k, m$ denotes positive integer with $a + mh = b$.

4. EXAMPLES

In this section, we provide two examples to illustrate our main results. The first one concerns the birth-and-death system and illustrates the verification of the imposed condition in this paper. The second one is about the risk management problem and exhibits the usefulness of the value iteration algorithm for computing the value function and an optimal policy.

Example 4.1. (Optimal control of birth-and-death system; see Example 6.1 in [6]) Consider a controlled birth-and-death system (33 – 34) below, in which the state variable represents the population size. The positive constants ρ and μ denote natural birth and death rates, respectively. In state 0, the decision maker selects an action a from a finite set $A(0)$. This action may increase ($u_2(0,a) \geq 0$) or decrease ($u_2(0,a) \leq 0$) the immigration parameter. In state $i \in \{1,2,3,\ldots\}$, the decision maker takes an action $a \in A(i)$, where $A(i)$ is a finite set. This action may increase ($u_2(i,a) \geq 0$) or decrease ($u_2(i,a) \leq 0$) the immigration parameter, and also increase ($u_1(i,a) \geq 0$) or decrease ($u_1(i,a) \leq 0$) the emigration parameter. Moreover, the decision maker takes a action $a \in A(i), i \in E$, which incurs a loss at the loss rate $c(i,a) \geq 0$.

This birth-and-death system can be described by a continuous-time Markov decision process. Suppose that the corresponding transition rates are given by

For $i = 0$ and $a \in A(0)$,

$$q(1|0,a) := -q(0|0,a) = u_2(0,a), \quad q(j|0,a) = 0 \quad for \quad j \geq 2. \qquad (33)$$

For $i \geq 1$ and $a \in A(i)$

$$q(j|i,a) = \begin{cases} \mu i + u_1(i,a), & \text{if } j = i-1, \\ -(\rho+\mu)i - u_1(i,a) - u_2(i,a), & \text{if } j = i, \\ \rho i + u_2(i,a), & \text{if } j = i+1, \\ 0, & \text{otherwise.} \end{cases} \tag{34}$$

The goal here is to give some suitable conditions ensuring the existence of an optimal policy. To do so, we establish the following conditions.

B1. $\mu i + u_1(i,a) \geq 0$ and $\rho i + u_2(i,a) \geq 0$ for all $a \in A(i)$ and $i \geq 1$; and $u_2(0,a) \geq 0$ for all $a \in A(0)$.

B2. $\|u_k\| := \sup_{(i,a)\in K} |u_k(i,a)| < \infty$ for $k = 1, 2$.

Under these conditions, we obtain the following :

Proposition 4.1. Under the conditions **B1**, **B2**, the birth-and-death system satisfies Assumption 2.1 and 3.1. Thus, in particular, the existence of a risk probability optimal policy is ensured by Theorem 3.6.

P r o o f . To verify Assumption 2.1, set $V(i) := i+1$ for $i \geq 0$, $M_0 := \rho+\mu+\|u_1\|+\|u_2\|$. Since conditions **B1** and **B2** are satisfied, we have

$$q^*(i) = \sup_{a\in A(i)} q_i(a) \leq M_0 V(i) \tag{35}$$

for all $i \in E$, which implies that Assumption 2.1(b) holds.

Moreover, using (33) and (34), we obtain, for $a \in A(0)$

$$\sum_{j\in E} V(j)q(j|0,a) = u_2(0,a) \leq (\rho+\mu)V(0) + M_0, \tag{36}$$

for $i \geq 1$ and $a \in A(i)$. We also obtain

$$\sum_{j\in E} V(j)q(j|i,a) = (\rho-\mu)i - u_1(i,a) + u_2(i,a) \leq (\rho+\mu)V(i) + M_0, \tag{37}$$

which together with (35), (36) imply that Assumption 2.1 holds with $c_0 := \rho+\mu$ and $b_0 := M_0$.

It follows from the finiteness of $A(i)$ and Remark 3.1 that Assumption 3.1 holds. Then, by Theorem 3.6, we know that the risk probability optimal policy exists. □

Example 4.2. (A risk management problem) Consider a startup company with three running status 0, 1 and 2, where the state 0 represents the company goes bankrupt, the state 1 represents the company is running normally, the state 2 represents the company has been a very good cash generator. In state 0, the company went bankrupt and could not pay any losses, which means that the decision maker do not need to choose any

decision action (we denote by a_{01}) and $c(0, a_{01}) = 0, a_{01} \in A(0)$. In state 1, the decision maker can choose a financing way a_{11} incurring in a loss rate $c(1, a_{11}) \geq 0$ or another financing way a_{12} incurring in a loss rate $c(1, a_{12}) \geq 0$. In state 2, the decision maker can choose a high yield financing way a_{21} incurring in a higher loss rate $c(2, a_{21}) \geq 0$ or an ordinary financing way a_{22} incurring in a lower loss rate $c(2, a_{22}) \geq 0$. The evolution of this controlled system as follows: when the system is in the state $i \in \{1, 2\}$, and the action $a \in A(i) = \{a_{i1}, a_{i2}\}$ is chosen, the system remains at i for an exponential-distributed random time with the parameter $q(i|i, a)$, and then moves to a new state j with probability $\frac{q(j|i,a)}{q_i(a)}(q_i(a) \neq 0, j = 0, 1, 2)$. For this system, the main objective of the decision marker is to find an optimal policy for the risk probability with loss rate during a fixed finite horizon $[0, T]$.

From the above evolution, this controlled system can be regarded as a model of CTMDPs with the state space $E = \{0, 1, 2\}$, the action sets $A(0) = \{a_{01}\}, A(1) = \{a_{11}, a_{12}\}, A(2) = \{a_{21}, a_{22}\}$. Moreover, we assume that the planning horizon $T = 15$, the transition rates are given by

$$\begin{array}{lll} q(0|0, a_{01}) = 0, & q(1|0, a_{01}) = 0, & q(2|0, a_{01}) = 0, \\ q(0|1, a_{11}) = 0.0144, & q(1|1, a_{11}) = -0.18, & q(2|1, a_{11}) = 0.1656, \\ q(0|1, a_{12}) = 0.0072, & q(1|1, a_{12}) = -0.12, & q(2|1, a_{12}) = 0.1128, \\ q(0|2, a_{21}) = 0.006, & q(1|2, a_{21}) = 0.294, & q(2|2, a_{21}) = -0.3, \\ q(0|2, a_{22}) = 0.0018, & q(1|2, a_{22}) = 0.0582, & q(2|2, a_{22}) = -0.06, \end{array} \tag{38}$$

and the loss rates are given by

$$c(0, a_{01}) = 0, \quad c(1, a_{11}) = 4, \quad c(1, a_{12}) = 3, \quad c(2, a_{21}) = 5, \quad c(2, a_{22}) = 2.$$

From (38), we know that the transition rates are uniformly bounded, the state space and the action space are finite. Then, by Remark 2.3 and 3.1, we obtain that Assumption 2.1 and 3.1 are satisfied under the condition $V \equiv 1$ in Lemma 2.2. Thus, we can use the value iteration algorithm in Theorem 3.6 to compute the value function and optimal policies.

For $i = 0$, by (38), we know the state 0 is a absorbing state. Hence, from $c(0, a_{01}) = 0$, we have the value function $U^*(0, \lambda, t) = U^\pi(0, \lambda, t) = 0$ for any policy $\pi \in \Pi$.

For $i = 1, 2, \lambda \in [0, +\infty)$ and $t \in [0, 15]$, by Theorem 3.6(a), we compute the function $U^*(2, \lambda, t)$ as follows:

Step 1: Set $U^*_{-1}(i, \lambda, t) := 0$.

Step 2: Employing (30) and (31), we calculate the functions $U^a_{n+1}(i, \lambda, t)$ and $U^*_{n+1}(i, \lambda, t)$ as follows:

For $i = 1, a \in A(1), n \geq 1$,

$$\begin{aligned} H^{a_{11}} U^*_n(1, \lambda, t) &= I_{(\lambda, +\infty)}(4t) e^{-0.18t} \\ &\quad + 0.08 \times 0.18 \times \int_0^t U^*_n(0, \lambda - 4u, t - u) e^{-0.18u} \, \mathrm{d}u \\ &\quad + 0.92 \times 0.18 \times \int_0^t U^*_n(2, \lambda - 4u, t - u) e^{-0.18u} \, \mathrm{d}u, \end{aligned}$$

$$
\begin{aligned}
H^{a_{12}}U_n^*(1,\lambda,t) &= I_{(\lambda,+\infty)}(3t)e^{-0.12t} \\
&\quad +0.06\times 0.12\times \int_0^t U_n^*(0,\lambda-3u,t-u)e^{-0.12u}\,\mathrm{d}u \\
&\quad +0.94\times 0.12\times \int_0^t U_n^*(2,\lambda-3u,t-u)e^{-0.12u}\,\mathrm{d}u, \\
U_{n+1}^*(1,\lambda,t) &= \min\{H^{a_{11}}U_n^*(1,\lambda,t), H^{a_{12}}H_n^*(1,\lambda,t)\}.
\end{aligned}
$$

For $i=2, a\in A(2), n\geq 1$,

$$
\begin{aligned}
H^{a_{21}}U_n^*(2,\lambda,t) &= I_{(\lambda,+\infty)}(5t)e^{-0.3t} \\
&\quad +0.02\times 0.3\times \int_0^t U_n^*(0,\lambda-5u,t-u)e^{-0.3u}\,\mathrm{d}u \\
&\quad +0.98\times 0.3\times \int_0^t U_n^*(1,\lambda-5u,t-u)e^{-0.3u}\,\mathrm{d}u, \\
H^{a_{22}}U_n^*(2,\lambda,t) &= I_{(\lambda,+\infty)}(2t)e^{-0.06t} \\
&\quad +0.03\times 0.06\times \int_0^t U_n^*(0,\lambda-2u,t-u)e^{-0.06u}\,\mathrm{d}u \\
&\quad +0.97\times 0.06\times \int_0^t U_n^*(1,\lambda-2u,t-u)e^{-0.06u}\,\mathrm{d}u, \\
U_{n+1}^*(2,\lambda,t) &= \min\{H^{a_{21}}U_n^*(2,\lambda,t), H^{a_{22}}U_n^*(2,\lambda,t)\}.
\end{aligned}
$$

Step 3: For each $i=1,2$, if $|U_{n+1}^*(i,\lambda,t)-U_n^*(i,\lambda,t)|<10^{-12}$, the iteration stops. Then, go to step 4, the value U_{n+1}^* is usually received as U^*; or else, go black to step 2 and by replacing n with $n+1$.

Step 4: For each $i=1,2, t=10,15$, drawing the graphs of these functions $H^aU^*(i,\lambda,t)$, $U^*(i,\lambda,t)$, see Figures 1 – 4.

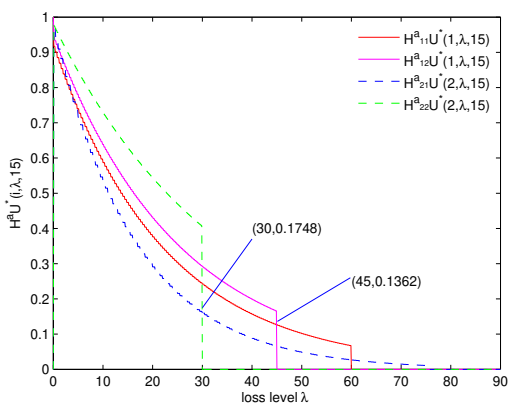

Fig. 1. The function $H^aU^*(i,\lambda,15)$.

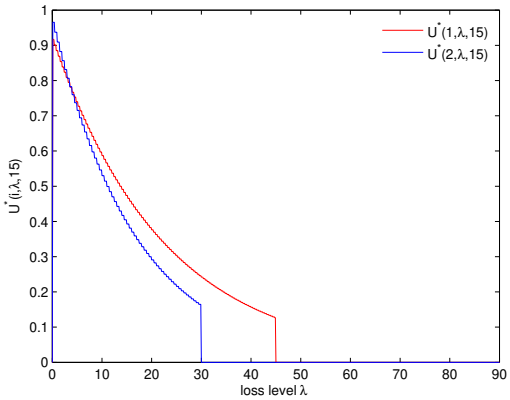

Fig. 2. The value function $U^*(i, \lambda, 15)$.

From Figures 1 – 4, we can observe the following conclusions.
(a) From Figures 1 – 2, we see that at time $s_0 = 0$, the planning horizon is $t = 15$ and $U^*(i, \lambda, 15) = H^{a_{i1}}U^*(i, \lambda, 15) = H^{a_{i2}}U^*(i, \lambda, 15) = 0$ with $i = 1, 2, \lambda \geq 90$. At state 1, when $0 < \lambda < 45$, $H^{a_{11}}U^*(1, \lambda, 15)$ is less than $H^{a_{12}}U^*(1, \lambda, 15)$; when $45 \leq \lambda < 90$, $H^{a_{12}}U^*(1, \lambda, 15)$ is less than $H^{a_{11}}U^*(1, \lambda, 15)$. At state 2, when $0 < \lambda < 30$, $H^{a_{21}}U^*(2, \lambda, 15)$ is less than $H^{a_{22}}U^*(1, \lambda, 15)$; when $30 \leq \lambda < 90$, $H^{a_{22}}U^*(2, \lambda, 15)$ is less than $H^{a_{21}}U^*(2, \lambda, 15)$. In this case, under the minimum risk probability criterion the optimal action is selected according to the following formula:

$$f^*(1, \lambda, 15) = \begin{cases} a_{11}, & 0 \leq \lambda < 45; \\ a_{12}, & 45 \leq \lambda < 90; \\ a_{11} = a_{12}, & \lambda \geq 90. \end{cases} \quad f^*(2, \lambda, 15) = \begin{cases} a_{21}, & 0 \leq \lambda < 30; \\ a_{22}, & 30 \leq \lambda < 90; \\ a_{21} = a_{22}, & \lambda \geq 90. \end{cases} \tag{39}$$

(b) From Figures 3 – 4, we see that $U^*(i, \lambda, 10) = H^{a_{i1}}U^*(i, \lambda, 10) = H^{a_{i2}}U^*(i, \lambda, 10) = 0$ with $i = 1, 2, \lambda \geq 90$. At state 1, when $0 < \lambda < 30$, $H^{a_{11}}U^*(1, \lambda, 10)$ is less than $H^{a_{12}}U^*(1, \lambda, 10)$; when $30 \leq \lambda < 90$, $H^{a_{12}}U^*(1, \lambda, 10)$ is less than $H^{a_{11}}U^*(1, \lambda, 10)$. At state 2, when $0 < \lambda < 20$, $H^{a_{21}}U^*(2, \lambda, 10)$ is less than $H^{a_{22}}U^*(1, \lambda, 10)$; when $20 \leq \lambda < 90$, $H^{a_{22}}U^*(2, \lambda, 10)$ is less than $H^{a_{21}}U^*(2, \lambda, 10)$. In this case, under the minimum risk probability criterion the optimal action is selected according to the following formula:

$$f^*(1, \lambda, 10) = \begin{cases} a_{11}, & 0 \leq \lambda < 30; \\ a_{12}, & 30 \leq \lambda < 90; \\ a_{11} = a_{12}, & \lambda \geq 90. \end{cases} \quad f^*(2, \lambda, 15) = \begin{cases} a_{21}, & 0 \leq \lambda < 20; \\ a_{22}, & 20 \leq \lambda < 90; \\ a_{21} = a_{22}, & \lambda \geq 90. \end{cases} \tag{40}$$

It follows from (a), we know that at the initial time $s_0 = 0$, when the system state is $i_0 \in \{1, 2\}$, the loss level $\lambda_0 > 0$ and the planning horizon $t_0 = 15$, the controller chooses an action $\tilde{f}_0(i_0, \lambda_0, t_0) := f^*(i_0, \lambda_0, t_0)$ according to (39). Once such an action is taken, the system stays at state i_0 until time s_1, at which point the system moves to a new state $i_1 \in \{0, 1, 2\}(i_1 \neq i_0)$ and a loss $c(i_0, \tilde{f}_0(i_0, \lambda_0, t_0))\theta_1$ is incurred. If the system state

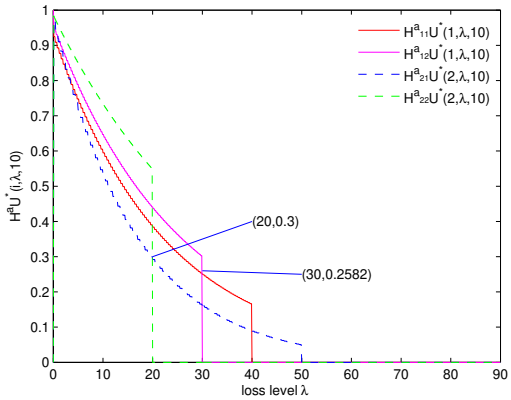

Fig. 3. The function $H^aU^*(i,\lambda,10)$.

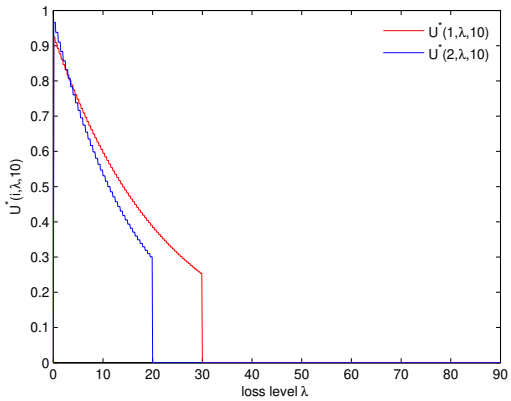

Fig. 4. The value function $U^*(2,\lambda,10)$.

$i_1 = 0$, the system will remain in state forever, as can be seen from (38). If the system state $i_1 \neq 0$, based on the current state i_1, loss level $\tilde{\lambda}_1 = L_1(i_0, \lambda_0, \tilde{f}_0(i_0, \lambda_0, t_0), \theta_1) = [\lambda_0 - c(i_0, \tilde{f}_0(i_0, \lambda_0, t_0))\theta_1]^+$, planning horizon $t_1 = L_2(t_0, \theta_1) = [t_0 - \theta_1]^+$ with $\theta_1 = s_1 - s_0$, the controller chooses another action $\tilde{f}_1(i_0, \lambda_0, t_0, s_1, i_1, \lambda_1, t_1) := f^*(i_1, \tilde{\lambda}_1, t_1) \in A(i_1)$, where the holding time $\theta_1 = s_1 - s_0$ satisfies the exponential distribution $1 - e^{-q_{i_0}(\tilde{f}_0(i_0,\lambda_0,t_0))\theta_1}$. For example, suppose that $\theta_1 = 5$, by (4) and (5), we know that the corresponding planning horizon $t_1 = 10$. Then, from (40), the optimal action $\tilde{f}_1$ is taken. The evolution of this system is repeated, which together with Theorem 3.6 (c) gives that $\pi^* = \{\tilde{f}_0, \tilde{f}_1, \ldots\}$ is optimal.

ACKNOWLEDGEMENT

This work was supported by National Natural Science Foundation of China (Grant No. 11961005, 11801590); PhD research startup foundation of Guangxi University of Science and Technology(Grant No.18Z06); Foundation of Guangxi Educational Committee(Grant No. KY2019YB0369); Guangxi Natural Science Foundation Program(Grant No.2020GXNSFAA297196).

(Received May 15, 2019)

REFERENCES

[1] K. Boda, J. A. Filar, and Y. L. Lin: Stochastic target hitting time and the problem of early retirement. IEEE Trans. Automat. Control *49* (2004), 409–419. DOI:10.1109/TAC.2004.824469

[2] M. Bouakiz and Y. Kebir: Target-level criterion in Markov decision process. J. Optim. Theory Appl. *86* (1995), 1–15. DOI:10.1007/BF02193458

[3] D. Bertsekas, S. Shreve: Stochastic Optimal Control: The Discrete-Time Case. Academic Press Inc, New York 1978.

[4] N. Bauerle and U. Rieder: Markov Decision Processes with Applications to Finance. Springer, Heidelberg 2011.

[5] E. Feinberg: Continuous time discounted jump Markov decision processes: a discrete-event approach. Math. Operat. Res. *29* (2004), 492–524. DOI:10.1287/moor.1040.0089

[6] X. P. Guo and O. Hernández-Lerma: Continuous-Time Markov Decision Process: Theorey and Applications. Springer-Verlag, Berlin 2009.

[7] X. P. Guo and A. Piunovskiy: Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates. Math. Oper. Res. *36* (2011), 105–132. DOI:10.1287/moor.1100.0477

[8] X. P. Guo, X. X. Huang, and Y. H. Huang: Finite-horizon optimality for continuous-time Markov decision processs with unbounded transition rates. Adv. Appl. Prob. *47* (2015), 1064–1087.DOI:10.1239/aap/1449859800

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

[10] Y. H. Huang and X. P. Guo: Optimal risk probability for first passage models in Semi-Markov processes. J. Math. Anal. Appl. *359* (2009), 404–420. DOI:10.1016/j.jmaa.2009.05.058

[11] Y. H. Huang and X. P. Guo: First passage models for denumberable Semi-Markov processes with nonnegative discounted cost. Acta. Math. Appl. Sinica *27* (2011), 177–190. DOI:10.1007/s10255-011-0061-2

[12] Y. H. Huang, X. P. Guo, and Z. F. Li: Minimum risk probability for finite horizon semi-Markov decision process. J. Math. Anal. Appl. *402* (2013), 378–391. DOI:10.1016/j.jmaa.2013.01.021

[13] X. X. Huang, X. L. Zou, and X. P. Guo: A minimization problem of the risk probability in first passage semi-Markov decision processes with loss rates. Sci. China Math. *58* (2015), 1923–1938. DOI:10.1007/s11425-015-5029-x

[14] H. F. Huo, X. L. Zou, and X. P. Guo: The risk probability criterion for discounted continuous-time Markov decision processes. Discrete Event Dynamic system: Theory Appl. *27* (2017), 675–699. DOI:10.1007/s10626-017-0257-6

[15] H. F. Huo, X. Wen: First passage risk probability optimality for continuous time Markov decision processes. Kybernetika *55* (2019), 114–133. DOI:10.14736/kyb-2019-1-0114

[16] H. F. Huo, X.P. Guo: Risk probability minimization problems for continuous time Markov decision processes on finite horizon. IEEE trans. Automat. Control *65* (2020), 3199–3206. DOI:10.1109/TAC.2019.2947654

[17] J. Jacod: Multivariate point processes: Predictable projection, Radon–Nicodym derivatives, representation of martingales. Z. Wahrscheinlichkeitstheorie und verwandte Gebiete *31* (1975), 235–253. DOI:10.1007/BF00536010

[18] J. Janssen and R. Manca: Semi-Markov Risk Models For Finance, Insurance, and Reliability. Springer-Verlag, New York 2006.

[19] Q. L. Liu, X. L. Zou: A risk minimization problem for finite horizon semi-Markov decision processes with loss rates. J. Dynamics Games *5* (2018), 143–163. DOI:10.3934/jdg.2018009

[20] A. Piunovskiy and Y. Zhang: Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. SIAM J. Control Optim. *49* (2011), 2032–2061. DOI:10.1137/10081366x

[21] Y. Ohtsubo and K. Toyonaga: Optimal policy for minimizing risk models in Markov decision processes. J. Math. Anal. Appl. *271* (2002), 66–81. DOI:10.1016/S0022-247X(02)00097-5

[22] Y. Ohtsubo: Risk minimization in optimal stopping problem and applications. J. Oper. Res. Soc. Japan *46* (2003), 342–352. DOI:10.15807/jorsj.46.342

[23] Y. Ohtsubo and K. Toyonaga: Equivalence classes for optimizing risk models in Markov decision processes. Math. Methods Oper. Res. *60* (2004), 239–250. DOI:10.1007/s001860400361

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

[25] M. Sakaguchi and Y. Ohtsubo: Optimal threshold probability and expectation in semi-Markov decision processes. Appl. Math. Comput. *216* (2010), 2947–2958. DOI:10.1007/s001860400361

[26] M. J. Sobel: The variance of discounted Markov decision processes. J. Appl. Probab. *19* (1982), 744–802.

[27] Q. D. Wei and X. P. Guo: Constrained semi-Markov decision processes with ratio and time expected average criteria in Polish spaces. Optimization *64* (2015), 1593–1623. DOI:10.1080/02331934.2013.860686

[28] D. J. White: Minimizing a threshold probability in discounted Markov decision processes. J. Math. Anal. Appl. Optim. *173* (1993), 634–646. DOI:10.1006/jmaa.1993.1093

[29] C. B. Wu and Y. L. Lin: Minimizing risk models in Markov decision processes with policies depending on target values. J. Math. Anal. Appl. *231* (1999), 47–67. DOI:10.1006/jmaa.1998.6203

[30] R. Wu and K. Fang: A risk model with delay in claim settlement. Acta Math. Applic. Sinica *15* (1999), 352–360. DOI:/10.1007/BF02684035

[31] S. X. Yu, Y. L. Lin, and P. F. Yan: Optimization models for the first arrival target distribution function in discrete time. J. Math. Anal. Appl. *225* (1998), 193–223. DOI:10.1006/jmaa.1998.6015

[32] L. Xia: Optimization of Markov decision processes under the variance criterion Automatica *73* (2016), 269–278. DOI:10.1016/j.automatica.2016.06.018

Haifeng Huo, Corresponding author. School of Science, Guangxi University of Science and Technology, Liuzhou, 545006. P. R. China.
e-mail: xiaohuo08ok@163.com

Xian Wen, School of Science, Guangxi University of Science and Technology, Liuzhou, 545006. P. R. China.
e-mail: wenxian879@163.com