Xian Wen; Jinhua Cui; Haifeng Huo

The risk probability optimal problem for infinite discounted semi-Markov decision processes

Kybernetika, Vol. 61 (2025), No. 4, 447–466

Persistent URL: http://dml.cz/dmlcz/153068

KYBERNETIKA — VOLUME *61* (2025), NUMBER 4, PAGES 447–466

THE RISK PROBABILITY OPTIMAL PROBLEM FOR INFINITE DISCOUNTED SEMI-MARKOV DECISION PROCESSES

Xian Wen, Jinhua Cui and Haifeng Huo

This paper investigates the risk probability minimization problem for infinite horizon semi-Markov decision processes (SMDPs) with varying discount factors. First, we establish the standard regularity condition to guarantee the state process is non-explosive. Furthermore, based only on the non-explosion of the state process, we use value iteration technique to establish the optimality equation satisfied by the value function, and prove the uniqueness of the solution and the existence of the risk probability optimal policy. Our condition is weaker than the first arrival condition commonly used in existing literature. Finally, we develop a value iteration algorithm to compute the value function and optimal policy, and illustrate the feasibility and effectiveness of the algorithm through a numerical example.

Keywords: risk probability criterion, semi-Markov decision processes, value function, optimal policy, value iteration algorithm

Classification: 90C40, 60E20

1. INTRODUCTION

As is well known, the risk probability optimal problem for Markov decision processes (MDPs) is an important dynamic programming problem, that has been extensively researched and widely applied in various practical fields, including queueing systems[5, 15], finance [3], and communication networks [7]. According to the probability distribution characteristics of the holding time, the existing research can be classified into three main categories: (i) Discrete-time Markov decision processes (DTMDPs), where the sojourn time of the system state is a fixed constant; see, [19, 21, 23, 25, 26]. (ii) Continuous-time Markov decision processes (CTMDPs), where the sojourn time of the system state follows an exponential distribution; see, [13, 14, 24]. (iii) Semi-Markov decision processes (SMDPs), where the sojourn time of the system states follows an arbitrary probability distribution; see, [8, 9, 11, 12, 18]. The risk probability criterion in SMDPs are primarily focused on risk analysis of systems in finite-horizon. The corresponding dynamic behaviors for SMDPs in infinite horizon still not adequately investigated.

The main motivation of this paper is as follows:

DOI: 10.14736/kyb-2025-4-0447

(i) Compared to DTMDPs and CTMDPs, SMDPs are a more versatile stochastic dynamic programming model, capable of more flexibly modeling complex decision-making problems in the real world, such as wireless communication, queuing systems, and machine maintenance.

(ii) For infinite horizon MDPs, the existing research have mainly focused on the traditional expected criteria based on risk neutrality (see [4, 5, 7]), which optimize decision-making by characterizing the expected values of total rewards. However, these criteria fail to capture the risk characteristics of the system. Therefore, it is necessary to introduce a probability criterion that can measure risk by calculating the probability that the total rewards/costs exceeds a given threshold, providing a more precise method for risk assessment in stochastic system.

(iii) The discount factor, as a core parameter for converting future rewards/costs into present value, essentially represents the time value of money mathematically. Current research on infinite horizon risk probability CTMDPs [13] typically assumes a constant discount factor. However, in practical applications in finance and insurance [6, 15, 17, 26], non-constant discount factors are more common. Therefore, in this paper, we consider the varying discount factor to enable more precise and practical evaluation of the present value of future cash flows.

Compared with the classical expected optimal problem [2, 9, 10, 17, 20], the decision-maker considers not only the system states but also the additional cost levels in our model, thereby providing a more robust framework for risk decision-making. Based on these considerations, the classical expected utility theories, such as Dynkin's formula in [6] and the Banach fixed point theorem in [4], are no longer suitable for our model. To deal with the probability criterion for SMDPs, we need to redefine the history-dependent policies, construct a new probability measure. The contributions of this paper are as follows:

(i) Redefining policies: since cost levels are regarded as components of an extended state, as well as past states and decision epochs, we need to redefine history-dependent, Markov, and stationary policies respectively by implementing a k-component internal history; see Definition 1.

(ii) Reconstructing a probability space: Since the historical information includes additional cost levels, the theory of traditional expected MDPs cannot be directly applied to our model (see Ref. [2, 9, 10, 17, 20]). Therefore, for any redefined policy, initial state and cost levels, we reconstruct a new probability measure and state process by extending the Ionescu Tulcea theorem, as shown in (8) – (10).

(iii) Establishing the optimality equation: In contrast to the classical approaches using Dynkin's formula and the Feynman-Kac formula for expected optimality problem, our method is based on the iteration technique. As an advantage of developing this approach, we first establish the optimality equation, and prove that the value function is the unique solution to the corresponding optimality equation, see Theorem 3.

(iv) Proving the existence of an optimal policy: we establish a NEW fact (Theorem 1) using only the non-explosive condition of the state process to prove the existence of an optimal policy. However, the additional first passage condition in [8, 10, 12, 13] is required for the existence of optimal policies. Then, our condition is weaker than the previous work in [8, 10, 12, 13] for risk probability MDPs. Moreover, we can introduce

standard continuity-compactness condition and successfully extend the results of this paper to general action space without without any technical difficulty.

(v) Providing the value iteration algorithm: by utilizing Theorem 2, we introduce the value iteration algorithm to calculate the value function and the optimal policy. Finally, we explain and illustrate our main results through a specific example, and demonstrate the computation of the value function and the optimal policy.

The remainder of this paper is organized as follows. In Section 2, we introduce the optimal risk probability criterion for infinite discounted SMDPs. In Section 3, we present the main results, including the establishment of the optimality equation and the proof of the existence of optimal policies. In Section 4, we illustrate our main results through a numerical example.

2. THE CONTROL MODEL

The model is composed of as follows:

$$\{E, A, (A(x), x \in E), Q(\cdot, \cdot|x, a), c(x, a), \alpha(x)\}, \tag{1}$$

where

- E represents a Borel state space, endowed with the Borel σ-algebras $\mathcal{B}(E)$.
- A represents a Borel action space, endowed with the Borel σ-algebras $\mathcal{B}(A)$. $A(x) \in \mathcal{B}(A)$ represents the set of permissible actions in the state $x \in E$, which is assumed to be finite. $K := \{(x, a)|x \in E, a \in A(x)\}$ denotes the set of all feasible pairs of states and actions.
- $Q(\cdot, \cdot|x, a)$ represents a semi-Markov kernel on $R^+ \times S$ given K, where $R^+ := [0, \infty)$. For any $u \in R^+$ and $D \in \mathcal{B}(E)$, when the action $a \in A(x)$ is selected at state x, the semi-Markov kernel $Q(u, D|x, a)$ represents the joint probability that the system's holding time is no more than $u \in R^+$ and the state x transfers into the set D. The semi-Markov kernel $Q(\cdot, \cdot|x, a)$ satisfies the following properties for $(x, a) \in K$:
  - (i) For any $D \in \mathcal{B}(E)$, $Q(\cdot, D|x, a)$ represents a non-decreasing, right-continuous function from R^+ to $[0, 1]$ with $Q(0, D|x, a) = 0$.
  - (ii) For any $u \in R^+$, $Q(u, \cdot|x, a)$ represents a sub-stochastic kernel on E.
  - (iii) For any $D \in \mathcal{B}(E)$, $P(D|x, a)$ represents a stochastic kernel on E, where $P(D|x, a) := \lim_{u\to\infty} Q(u, D|x, a)$.
- $c(x, a)$ denotes the cost rate, which is assumed to be a nonnegative measurable function on K.
- $\alpha(x)$ represents the discount factor, which depends on the state $x \in E$.

The infinite discounted risk probability SMDPs (1) evolves as follows: At the initial time $s_0 = 0$, the system has the initial state $x_0 \in E$ and the cost level (goal) $\lambda_0 \in R^+$. The cost level indicates that the decision maker will make his or her best effort to ensure that the total costs does not exceed the cost goal(level). Based on the initial state

information $h_0 := (x_0, \lambda_0)$, the decision maker can choose an action a_0 from the set of available actions $A(x_0)$. The system will change in the following two aspects: (i) The system state remains at the state x_0 until the moment s_1, at which point the system state transitions to a new state x_1 with the transition probability $Q(s_1, \mathrm{d}x_1|x_0, a_0)$. (ii) During the interval $[0, s_1]$, the system incurs the operational cost $\int_0^{s_1} e^{-\int_0^s \alpha(x_0)\mathrm{d}t} c(x_0, a_0)\mathrm{d}s$. To account for the time value, we introduce the state-dependent discount factor. At the moment s_1, the cost level becomes $\lambda_1 := e^{\int_0^{s_1} \alpha(x_0)\mathrm{d}t}(\lambda_0 - \int_0^{s_1} e^{-\int_0^s \alpha(x_0))\mathrm{d}t} c(x_0, a_0)\mathrm{d}s)$. At the new decision making moment, the decision maker selects a new action $a_1 \in A(x_1)$ based on historical information $h_1 := (x_0, \lambda_0, a_0, s_1, x_1, \lambda_1)$. The system evolves continuously through (i) and (ii). At the decision-making moment $s_k, k = 2, 3, \ldots$, the decision maker chooses a new action based on the system's historical information

$$h_k := (x_0, \lambda_0, a_0, s_1, x_1, \lambda_1, a_1, \ldots, s_k, x_k, \lambda_k), \tag{2}$$

where s_k represents the kth decision epoch; x_k represents the state of the system during the interval $[s_k, s_{k+1})$; a_k denotes the action chosen by the decision maker at epoch s_k, $\theta_{k+1} := s_{k+1} - s_k$ indicates the sojourn time at the state x_k; λ_k represents the cost level, and its variation pattern satisfies the following equation

$$\begin{aligned} \lambda_{k+1} &:= L(x_k, \lambda_k, a_k, \theta_{k+1}) \\ &:= e^{\int_0^{\theta_{k+1}} \alpha(x_k)\mathrm{d}t}(\lambda_k - \int_0^{\theta_{k+1}} e^{-\int_0^s \alpha(x_k)\mathrm{d}t} c(x_k, a_k)\mathrm{d}s), \end{aligned} \tag{3}$$

for $\lambda_0 = \hat{\lambda}_0$. The set of all admissible histories h_k is defined as $H_0 := E \times R$ and $H_k := (E \times R \times A \times (0, +\infty])^k \times E \times R$. During the evolution process, the system needs to pay a series of costs. Based on the risk probability criterion, the decision maker considers how to choose the optimal action to ensure that the risk of system operation is optimized.

Remark 1. For the risk probability optimization problem, the decision maker not only considers the usual system state when making decisions but also takes into account the cost levels. This is a fundamental difference from classical expected optimization problem.

Definition 1. *A randomized history-dependent* policy $\pi = \{\pi_k, k \geq 0\}$ represents a sequence of stochastic kernel π_k on A given H_k with

$$\pi_k(A(x_k)|h_k) = 1 \quad \forall h_k \in H_k, k = 0, 1, \ldots.$$

The set of all randomized history-dependent policies is represented as Π.

Let Φ be the set of all stochastic kernels φ on A given S such that $\varphi(A(x)|x) = 1$. $\mathbb{F}$ denotes the set of all functions f such that $f(x, \lambda) \in A$ for $(x, \lambda) \in E \times R$.

Definition 2. A policy $\pi = \{\pi_k\} \in \Pi$ is called *randomized Markov* if there is a stochastic kernel $\varphi_k \in \varphi$ such that $\pi_k(\cdot|h_k) = \varphi_k(\cdot|x_k, \lambda_k)$ for each $h_k \in H_k$ and $k \geq 0$. Such a randomized Markov policy is denoted by $\pi = \{\varphi_k\}$.

A randomized Markov policy $\pi = \{\varphi_k\}$ is called deterministic Markov if there exists a measurable function $f_k \in F$ such that $\varphi_k(\cdot|x_k, \lambda_k)$ is a Dirac measure concentrated at $f_k(\cdot|x_k, \lambda_k)$ on $E \times R$.

A deterministic Markov policy $\pi = \{f_k\}$ is said to be *deterministic stationary* if the measurable function f_k is independent of k. For simplicity, such a policy is denoted by f. The set of all stationary policies is denoted by $\mathbb{F}$.

Let Π_{RM}, Π_{DM} and Π_{DS} be the class of all randomized Markov, deterministic Markov and deterministic stationary policies, respectively. Thus, $F = \Pi_{DS} \subset \Pi_{RS} \subset \Pi$.

Since the system state includes the additional cost levels, we need to construct a new probability space. The probability space is constructed as follows: The sample space is defined by $\Omega := \{(x_0, \lambda_0, a_0, s_1, x_1, \ \lambda_1, a_1, \ldots, s_k, \ x_k, \lambda_k, a_k, \ldots)|\ x_0 \in E, \lambda_0 \in R, a_0 \in A(x_0), s_l \in (0, \infty), x_l \in E, \lambda_l \in R, a_l \in A(x_l)$, for each $1 \leq l \leq k, k \geq 1\}$. The Borel σ-algebra of Ω is expressed as $\mathcal{F}$. For any $\omega := (x_0, \lambda_0, a_0, s_1, x_1, \lambda_1, a_1, \ldots, s_k, x_k, \lambda_k, a_k, \ldots) \in \Omega, k \geq 0$, the following random variables S_k, X_k, A_k, Λ_k on $(\Omega, \mathcal{F})$ are defined as follows:

$$\begin{aligned} &S_0(\omega) := s_0, S_{k+1}(\omega) := s_{k+1}, X_k(\omega) := x_k, \\ &A_k(\omega) := a_k, \Lambda_k = \lambda_k, S_\infty(\omega) := \lim_{k\to\infty} S_k(\omega). \end{aligned} \tag{4}$$

The corresponding state process $\{X_s, s \geq 0\}$ and the action process $\{A_s, s \geq 0\}$ are defined by

$$X_s(\omega) := \sum_{k\geq 0} I_{\{S_k(\omega)\leq s<S_{k+1}(\omega)\}} X_k(\omega) + \Delta I_{\{s\geq S_\infty(\omega)\}}, \tag{5}$$

$$A_s(\omega) := \sum_{k\geq 0} I_{\{S_k(\omega)\leq s<S_{k+1}(\omega)\}} A_k(\omega) + a_\Delta I_{\{s\geq S_\infty(\omega)\}}, \tag{6}$$

where $I_D(\cdot)$ represents the indicator function on the set D, $\Delta \notin E$ denotes a cemetery state, and a_Δ denotes an isolated point. The ω will be omitted for convenience.

For any $\pi \in \Pi, (x, \lambda) \in E \times R$, by the well-known Ionescu Tulcea theorem (e. g., Proposition 7.45 in [1]), there exists a unique probability space $(\Omega, \mathscr{F}, P^\pi_{(x,\lambda)})$ and the stochastic processes $\{X_s, A_s, s \geq 0\}$ satisfy the following properties: for any $u > 0, C \subset R, D \in \mathscr{B}(E), G \in \mathscr{B}(A), h_k \in H_k, k = 0, 1, \ldots$

$$P^\pi_{(x,\lambda)}(S_0 = 0, X_0 = x, \Lambda_0 = \lambda) = 1, \tag{7}$$

$$P^\pi_{(x,\lambda)}(A_k \in G|h_k) = \int_G \pi_k(\mathrm{d}a|h_k), \tag{8}$$

$$P^\pi_{(x,\lambda)}(S_{k+1} - S_k \leq u, X_{k+1} \in D|h_k, a_k) = \int_D Q(u, \mathrm{d}y|x_k, a_k), \tag{9}$$

$$P^\pi_{(x,\lambda)}(\Lambda_{k+1} \in C|h_k, a_k, s_{k+1}) = \int_C \delta_{L(x_k,\lambda_k,a_k,\theta_{k+1})}(\mathrm{d}\lambda_{k+1}), \tag{10}$$

where $\delta_{\lambda^*}(\lambda)$ denotes the Dirac measure on R at the point λ^*. The expectation operator is denoted by $\mathbb{E}^\pi_{(x,\lambda)}$ corresponding to the probability measure $P^\pi_{(x,\lambda)}$.

Definition 3. For each $(x,\lambda) \in E \times R, \pi \in \Pi$, the infinite discounted risk probability criterion is defined by

$$F^{\pi}(x,\lambda) := P^{\pi}_{(x,\lambda)}\Big(\int_0^{+\infty} e^{-\int_0^s \alpha(X_u)\mathrm{d}u} c(X_s, A_s)\mathrm{d}s \le \lambda \Big), \tag{11}$$

which represents the probability that the total costs do not exceed the cost goal under the policy π.

Definition 4. The infinite discounted risk probability value function is defined by

$$F^*(x,\lambda) = \inf_{\pi\in\Pi} F^{\pi}(x,\lambda) \quad \text{for all} \ \ (x,\lambda) \in E \times R. \tag{12}$$

A policy $\pi^* \in \Pi$ is called the risk probability optimal, if

$$F^{\pi^*}(x,\lambda) = F^*(x,\lambda) \quad \text{for all} \ \ (x,\lambda) \in E \times R. \tag{13}$$

The goal of this paper is to present the infinite discounted risk probability optimization problem, establish the optimality equation that the value function satisfies, prove the existence of an optimal policy, and provide a value iteration algorithm.

3. MAIN RESULTS

Let $\mathcal{F}_m$ be the set of all Borel measurable functions $F(\cdot,\cdot) : E \times R \to [0,1]$. For any $F \in \mathcal{F}_m, x \in E, a \in A(x)$ and $\varphi \in \Phi$, the operators T^{φ} and T are defined on $\mathcal{F}_m$ as follows: if $\lambda \ge 0$,

$$\begin{aligned} T^a F(x,\lambda) &:= \int_E \int_0^{+\infty} F\Big(y, e^{\alpha(x)u}(\lambda - \frac{c(x,a)(1-e^{-\alpha(x)u})}{\alpha(x)})\Big) Q(\mathrm{d}u, \mathrm{d}y|x,a), \\ T^{\varphi} F(x,\lambda) &:= \int_{A(x)} \varphi(\mathrm{d}a|x,\lambda) T^a F(x,\lambda), \qquad (14) \\ TF(x,\lambda) &:= \min_{a\in A(x)} T^a F(x,\lambda). \qquad (15) \end{aligned}$$

If $\lambda < 0$,

$$T^{\varphi} F(x,\lambda) = T^a F(x,\lambda) = TF(x,\lambda) := 0. \tag{16}$$

Similarly, the operators $(T^f)^n, T^n, n \ge 1$ on $\mathcal{F}_m$ are defined by $(T^f)^{n+1}F(x,\lambda) = T^f((T^f)^n F(x,\lambda))$ and $T^{n+1}F(x,\lambda) = T(T^n F(x,\lambda))$.

Let $\widetilde{\mathcal{F}}_m$ be the set of all Borel measurable functions $\widetilde{F} : E \times R \to [-1,1]$. For any $\widetilde{F} \in \widetilde{\mathcal{F}}_m, a \in A(x)$ and $\varphi \in \Phi$, the operators $(\widetilde{T}^{\varphi})^n$ are defined for $n \ge 1$. If $\lambda \ge 0$,

$$\begin{aligned} \widetilde{T}^a \widetilde{F}(x,\lambda) &:= \int_E \int_0^{+\infty} \widetilde{F}\Big(y, e^{\alpha(x)u}(\lambda - \frac{c(x,a)(1-e^{-\alpha(x)u})}{\alpha(x)})\Big) Q(\mathrm{d}u, \mathrm{d}y|x,a), \\ \widetilde{T}^{\varphi} F(x,\lambda) &:= \int_{A(x)} \varphi(\mathrm{d}a|x,\lambda) \widetilde{T}^a F(x,\lambda), \qquad (17) \end{aligned}$$

$$(\widetilde{T}^{\varphi})^{n+1}\widetilde{F}(x,\lambda) \quad := \quad T^{\varphi}((T^{\varphi})^{n}\widetilde{F}(x,\lambda)). \tag{18}$$

If $\lambda < 0$,

$$(\widetilde{T}^{\varphi})^{n}\widetilde{F}(x,\lambda) := 0. \tag{19}$$

Lemma 1. For any $(x,\lambda) \in E \times R, a \in A(x), F, G \in \mathcal{F}_m$, the operator T satisfies the following properties:

(a) If $F \geq G$, then $T^a F(x,\lambda) \geq T^a G(x,\lambda)$ and $TF(x,\lambda) \geq TG(x,\lambda)$.

(b) There exists a $f \in F$ such that $TF(x,\lambda) = T^f F(x,\lambda)$.

P r o o f. (a) For any $(x,\lambda) \in E\times R, a \in A(x), F, G \in \mathcal{F}_m, F \geq G$, based on the definition of T^a and T, it is easy to see that $T^a F(x,\lambda) \geq T^a G(x,\lambda)$ and $TF(x,\lambda) \geq TG(x,\lambda)$.

(b) For any $(x,\lambda) \in E \times R$, the finiteness of the set $A(x)$ which implies that the existence of $f \in \Pi_s$ in (b) is guaranteed. □

To ensure the existence of the optimal policy, we need to establish the following assumption to guarantee that the state process is non-explosive, where the non-explosion of state process means that the state process cannot have an infinite number of jumps within a finite time.

Assumption 1. For any $\pi \in \Pi, (x,\lambda) \in E \times R$, $P^{\pi}_{(x,\lambda)}(S_\infty = \infty) = 1$.

To verify Assumption 1, we give a sufficient condition in Lemma 2.

Lemma 2. If there are some constants δ, $\varepsilon > 0$ such that $Q(\delta, E|x,a) \leq 1-\varepsilon$ for each $(x,a) \in K$, then Assumption 1 holds.

P r o o f. It follows from Proposition 2.1 in [9]. □

Remark 2. It should be noted that Assumption 1 is used to ensure the existence of an optimal policy, see Theorem 3. Our condition is weaker than those existing in the research on infinite horizon [8, 10, 12, 13]. This is because, to ensure the existence of an optimal policy, the additional first passage condition must be established in existing research [4, 5, 13]. To facilitate the verification of Assumption 1, we provide the regularity condition in Lemma 2.

For any $(x,\lambda) \in E \times R$ and $\pi \in \Pi$, based on the non-explosion of the controlled state process, the continuity of the probability measure, $F^{\pi}(x,\lambda)$ can be reformulated as follows:

$$\begin{aligned} F^{\pi}(x,\lambda) &= P^{\pi}_{(x,\lambda)}\Big(\int_0^{+\infty} e^{-\int_0^s \alpha(X_t)\mathrm{d}t} c(X_s, A_s)\mathrm{d}s \leq \lambda\Big) \\ &= P^{\pi}_{(x,\lambda)}\Big(\sum_{m=0}^{\infty}\int_{S_m}^{S_{m+1}} e^{-\int_0^s \alpha(X_t)\mathrm{d}t} c(X_s, A_s)\mathrm{d}s \leq \lambda\Big) \end{aligned}$$

$$
\begin{aligned}
&= P^{\pi}_{(x,\lambda)}\Big(\bigcap_{n=1}^{\infty}\sum_{m=0}^{n}\int_{S_m}^{S_{m+1}} e^{-\int_0^s \alpha(X_t)\mathrm{d}t}c(X_s,A_s)\mathrm{d}s \le \lambda\Big)\\
&= \lim_{n\to\infty} P^{\pi}_{(x,\lambda)}\Big(\sum_{m=0}^{n}\int_{S_m}^{S_{m+1}} e^{-\int_0^s \alpha(X_t)\mathrm{d}t}c(X_s,A_s)\mathrm{d}s \le \lambda\Big)\\
&:= \lim_{n\to\infty} F^{\pi}_{n}(x,\lambda) \qquad (20)
\end{aligned}
$$

with $F^{\pi}_{-1}(x,\lambda) := I_{[0,+\infty)}(\lambda)$.

To prove the optimality equation satisfied by the value function, we first need to establish some lemmas.

Lemma 3. Under Assumption 1, the following conclusions hold.

(a) $F^{\pi}_{n} \in \mathcal{F}_m, n \ge -1$ and $F^{\pi} \in \mathcal{F}_m$ for each $(x,\lambda) \in E \times R, \pi = \{\varphi_0, \varphi_1, \ldots\} \in \Pi$.

(b) $F^{\pi}_{n+1} = T^{\varphi_0}F^{\pi}_{n}$ and $F^{\pi} = T^{\varphi_0}F^{\pi}$ for each $(x,\lambda) \in E \times R, \pi = \{\varphi_0, \varphi_1, \ldots\} \in \Pi$, ${}^1\pi := \{\varphi_1, \varphi_2, \ldots\}$ denotes the 1-shift policy of π.

In particular, for each $f \in \Pi_s$, $F^{f}(x,\lambda) = T^{f}F^{f}(x,\lambda)$.

P r o o f. (a) For each $(x,\lambda) \in E \times R, \pi \in \Pi$, firstly, we prove part (a) by using mathematical induction. Obviously, $F^{\pi}_{-1} = I_{[0,+\infty)}(\lambda) \in \mathcal{F}_m$. When $n = k \ge -1$, assume that the fact (a) holds. For $n = k+1 \ge -1$, by the properties of conditional expectation and (8), we have

$$
\begin{aligned}
& F^{\pi}_{k+1}(x,\lambda)\\
= \;& P^{\pi}_{(x,\lambda)}\Big(\sum_{m=0}^{k+1}\int_{S_m}^{S_{m+1}} e^{-\int_0^s \alpha(X_t)\mathrm{d}t}c(X_s,A_s)\mathrm{d}s \le \lambda\Big)\\
= \;& E^{\pi}_{(x,\lambda)}[I_{\{\sum_{m=0}^{k+1}\int_{S_m}^{S_{m+1}} e^{-\int_0^s \alpha(X_t)\mathrm{d}t}c(X_s,A_s)\mathrm{d}s \le \lambda, S_1 < \infty\}}]\\
= \;& E^{\pi}_{(x,\lambda)}[E^{\pi}_{(x,\lambda)}[I_{\{\sum_{m=0}^{k+1}\int_{S_m}^{S_{m+1}} e^{-\int_0^s \alpha(X_t)\mathrm{d}t}c(X_s,A_s)\mathrm{d}s \le \lambda, S_1 < \infty\}}|X_{S_0},\Lambda_0,S_1,X_{S_1},\Lambda_1]]\\
= \;& \int_{A(x)}\varphi_0(\mathrm{d}a|x)\int_E\int_0^{+\infty} P^{\pi}_{(x,\lambda)}\Big(\sum_{m=1}^{k+1}\int_{S_m}^{S_{m+1}} e^{-\int_0^s \alpha(X_t)\mathrm{d}t}c(X_s,A_s)\mathrm{d}s\\
& \le \lambda - \int_0^u e^{-\int_0^s \alpha(x)\mathrm{d}t}c(x,a)\mathrm{d}s|X_{S_0}=x,\Lambda_0=\lambda,S_1=u,X_{S_1}=y,\\
& \Lambda_1 = e^{\int_0^u \alpha(x)\mathrm{d}t}(\lambda - \int_0^u e^{-\int_0^s \alpha(x)\mathrm{d}t}c(x,a)\mathrm{d}s)\Big)\times Q(\mathrm{d}u,\mathrm{d}y|x,a)\\
= \;& \int_{A(x)}\varphi_0(\mathrm{d}a|x)\int_E\int_0^{+\infty} P^{\pi}_{(x,\lambda)}\Big(\sum_{m=1}^{k+1}\int_{S_m-u}^{S_{m+1}-u} e^{-\int_0^{l+u} \alpha(X_t)\mathrm{d}t}c(X_{l+u},A_{l+u})\mathrm{d}l\\
& \le \lambda - \int_0^u e^{-\int_0^s \alpha(x)\mathrm{d}t}c(x,a)\mathrm{d}s|X_{S_0}=x,\Lambda_0=\lambda,S_1=u,X_{S_1}=y,
\end{aligned}
$$

$$
\begin{aligned}
&\Lambda_1 = e^{\alpha(x)u}(\lambda - \int_0^u e^{-\int_0^s \alpha(x)\mathrm{d}t} c(x,a)\mathrm{d}s)\Big) \times Q(\mathrm{d}u, \mathrm{d}y|x,a)\\
= &\int_{A(x)} \varphi_0(\mathrm{d}a|x) \int_E \int_0^{+\infty} P^{1}{}_{(y, e^{\alpha(x)u}(\lambda - \frac{c(x,a)(1-e^{-\alpha(x)u})}{\alpha(x)}))}^{\pi} \Big(\sum_{m=0}^{k} \int_{S_m}^{S_{m+1}} e^{-\int_0^s \alpha(x)\mathrm{d}t}\\
&c(X_s, A_s)\mathrm{d}s \le e^{\alpha(x)u}(\lambda - \frac{c(x,a)(1-e^{-\alpha(x)u})}{\alpha(x)})\Big) Q(\mathrm{d}u, \mathrm{d}y|x,a)\\
= &\int_{A(x)} \varphi_0(\mathrm{d}a|x) \int_E \int_0^{+\infty} F_k^{1}{}^{\pi}\Big(y, e^{\alpha(x)u}(\lambda - \frac{c(x,a)(1-e^{-\alpha(x)u})}{\alpha(x)})\Big) Q(\mathrm{d}u, \mathrm{d}y|x,a)\\
:= &\ T^{\varphi_0} F_k^{1}{}^{\pi}(x,\lambda).
\end{aligned}
$$

Therefore, $F_n^\pi \in \mathcal{F}_m$ holds by induction. Since the limit of a sequence of measurable functions is also measurable, we know that $\lim_{n\to\infty} F_n^\pi = F^\pi \in \mathcal{F}_m$.

(*b*) For each $(x,\lambda) \in E \times R, n \ge -1$, based on part (a), we know that $F_{n+1}^\pi(x,\lambda) = T^{\varphi_0} F_n^{1}{}^{\pi}(x,\lambda)$. Letting $n \to \infty$, by using the dominated convergence theorem, we have $F^\pi(x,\lambda) = T^{\varphi_0} F^{1}{}^{\pi}(x,\lambda)$. Then, if $\pi = f \in \Pi_s$, we have $F^f(x,\lambda) = T^f F^f(x,\lambda)$. □

Theorem 1. Suppose that Assumption 1 holds. Then

(a) For any $(x,\lambda) \in E \times R$, if $F, G \in \mathcal{F}_m$, $F - G \le \widetilde{T}^f(F-G)$, then $F \le G$.

(b) For any $(x,\lambda) \in E \times R$, F^f is the unique solution in $\mathcal{F}_m$ to the equation $F^f = T^f F^f$.

P r o o f. (a) For any $(x,\lambda) \in E \times R, f \in \Pi_s, F, G \in \mathcal{F}_m$, firstly, we prove the fact for $n = 1, 2, \ldots,$

$$(\widetilde{T}^f)^n (F-G)(x,\lambda) \le P_{(x,\lambda)}^f(S_n < \infty). \tag{21}$$

Since $F - G \in \widetilde{\mathcal{F}}_m$, by (17), we obtain

$$(\widetilde{T}^f)^n (F-G)(x,\lambda) \le P_{(x,\lambda)}^f(S_n < \infty). \tag{22}$$

Since $F - G \in \widetilde{\mathcal{F}}_m$, by (17), we obtain

$$
\begin{aligned}
&\widetilde{T}^f(F-G)(x,\lambda)\\
= &\int_E \int_0^{+\infty} (F-G)\Big(y, e^{\alpha(x)u}(\lambda - \frac{c(x,f)(1-e^{-\alpha(x)u})}{\alpha(x)})\Big) Q(\mathrm{d}u, \mathrm{d}y|x,f),\\
\le &\int_E \int_0^{+\infty} Q(\mathrm{d}u, \mathrm{d}y|x,f)\\
= &\ P_{(x,\lambda)}^f(S_1 < \infty).
\end{aligned}
$$

When $n = k$, assume that the fact (22) holds. According to the definition of the operator and the induction, we obtain

$$\begin{aligned}
& (\widetilde{T}^f)^{k+1}(F-G)(x,\lambda) \\
= \;& \widetilde{T}^f(\widetilde{T}^f)^k(F-G)(x,\lambda) \\
= \;& \int_E\int_0^{+\infty}(\widetilde{T}^f)^k(F-G)\Big(y, e^{\alpha(x)u}(\lambda-\frac{c(x,f)(1-e^{-\alpha(x)u})}{\alpha(x)})\Big)Q(\mathrm{d}u,\mathrm{d}y|x,f), \\
\leq \;& \int_E\int_0^{+\infty} P^f_{(y,e^{\alpha(x)u}(\lambda-\frac{c(x,f)(1-e^{-\alpha(x)u})}{\alpha(x)}))}(S_k<\infty)Q(\mathrm{d}u,\mathrm{d}y|x,f). \qquad (23)
\end{aligned}$$

On the other hand, by the properties of conditional expectation, we have

$$\begin{aligned}
& P^f_{(x,\lambda)}(S_{k+1}<\infty) \\
= \;& E^f_{(x,\lambda)}[E^f_{(x,\lambda)}[I_{\{S_{k+1}<\infty\}}|X_{S_0},\Lambda_0,S_1,X_{S_1},\Lambda_1]] \\
= \;& \int_E\int_0^{+\infty} P^f_{(x,\lambda)}\Big(S_{k+1}<\infty|X_{S_0}=x,\Lambda_0=\lambda,S_1=u, \\
& X_{S_1}=y,\Lambda_1=e^{\alpha(x)u}(\lambda-\frac{c(x,f)(1-e^{-\alpha(x)u})}{\alpha(x)})\Big)Q(\mathrm{d}u,\mathrm{d}y|x,f) \\
= \;& \int_E\int_0^{+\infty} P^f_{(y,e^{\alpha(x)u}(\lambda-\frac{c(x,f)(1-e^{-\alpha(x)u})}{\alpha(x)}))}(S_k<\infty)Q(\mathrm{d}u,\mathrm{d}y|x,f),
\end{aligned}$$

which combined with (23) shows that the inductive hypothesis holds. Thus, the fact (22) is proved.

Finally, by using (22) and the induction hypothesis, it follows that

$$\begin{aligned}
& F(x,\lambda)-G(x,\lambda) \\
\leq \;& \widetilde{T}^f\big(F(x,\lambda)-G(x,\lambda)\big) \\
\leq \;& (\widetilde{T}^f)^n\big(F-G\big)(x,\lambda) \\
\leq \;& P^f_{(x,\lambda)}\big(S_n<\infty\big)\ \forall\, n\geq 1. \qquad (24)
\end{aligned}$$

Letting $n\to\infty$, under Assumption 1, by the continuity of the probability measure, we have

$$F(x,\lambda)-G(x,\lambda)\leq P^f_{(x,\lambda)}(S_\infty<\infty)=0,$$

which implies that $F\leq G$.

(b) For any $(x,\lambda)\in E\times R, f\in\Pi_s$, based on Lemma 3, we know that $F^f=T^fF^f$. Suppose that $G^f\in\mathcal{F}_m$ is another solution to the equation $G^f=T^fG^f$. Then, $F^f-G^f=\widetilde{T}^f(F^f-G^f)$. Consequently, by part (a), we have $F^f=G^f$. □

Theorem 2. Suppose that Assumption 1 holds, for any $(x,\lambda)\in E\times R$, let $F^*_{-1}:=I_{[0,+\infty)}(\lambda), F^*_{n+1}:=TF^*_n, n\geq -1$. Then, $\lim_{n\to\infty}F^*_n=F^*$.

P r o o f. Since $F^*_{-1} := I_{[0,+\infty)}(\lambda), F^*_{n+1} := TF^*_n, n \geq -1$, by Lemma 1, we know that $0 \leq F^*_{n+1} \leq F^*_n \leq 1$ and $\lim_{n\to\infty} F^*_n := \widetilde{F}$ exists.

To prove $\tilde{F} \leq F^*$, We first use induction to prove the following fact:

$$F^*_n \leq F^\pi_n \quad \text{for all } n \geq -1, \pi \in \Pi. \tag{25}$$

For any $\pi \in \Pi$, since $F^*_{-1} = F^\pi_{-1} := I_{[0,+\infty)}(\lambda)$, when $n = -1$, the conclusion obviously holds. Assume that $F^*_k \leq F^\pi_k$ for $n = k \geq -1$. Then, for any $(x, \lambda) \in E \times R, \pi = \{\varphi_0, \varphi_1, \ldots\} \in \Pi$, we have

$$F^*_{k+1} = TF^*_k \leq TF^{{}^1\pi}_k \leq T^{\varphi_0} F^{{}^1\pi}_k = F^\pi_{k+1},$$

where the second inequality and the last equality follow from the induction hypothesis and Lemma 3(b), respectively. Hence, the inductive hypothesis is satisfied, and the conclusion has been proven. Letting $n \to \infty$ in (25), by (20), we have $\tilde{F} \leq F^*$.

To prove $\tilde{F} \geq F^*$, for any $(x, \lambda) \in E \times R$ and $n \geq -1$, firstly, we prove that there exists a policy $\eta \in \Pi_{RM}$ such that $F^*_n = F^\eta_n$ by induction. Since $F^*_{-1} = F^\pi_{-1} = I_{[0,+\infty)}(\lambda)$ for any $\pi \in \Pi_{RM}$, this fact is true for $n = -1$. Suppose that there exists a policy $\eta \in \Pi_{RM}$ such that $F^*_k = F^\eta_k$ for $n = k \geq -1$. Moreover, by using Lemma 1(b), we know that there is a policy $f \in \Pi_s$ such that $T^f F^*_k = TF^*_k$. Then, letting $\zeta = \{f, \eta\} \in \Pi_{RM}$, by the induction hypothesis and Lemma 3(b), we have

$$F^*_{k+1} = TF^*_k = T^f F^*_k = T^f F^\eta_k = F^\zeta_{k+1}.$$

Then, the inductive hypothesis is valid. The conclusion is proven, that is, there exists a a policy $\eta \in \Pi_{RM}$ satisfies

$$F^*_n = F^\eta_n \geq F^\eta \geq F^*.$$

Hence, letting $n \to \infty$, according to (20), we can deduce that $\tilde{F} \geq F^*$, which implies $\tilde{F} = F^*$. □

Remark 3. For each $(x, \lambda) \in E \times R$, based on Theorem 2, we have established the so-called value iteration algorithm to calculate the value function as follows: Let $F^*_{-1} := I_{[0,+\infty)}(\lambda)$ and $F^*_{n+1} = TF^*_n, n \geq -1$. Then, $\lim_{n\to\infty} F^*_n = F^*$.

Theorem 3. Under Assumption 1, then

(a) F^* is the unique solution to the risk probability optimality equation $F^* = TF^*$.

(b) There is a policy $f^* \in \Pi_s$ such that $F^*(x, \lambda) = T^{f^*} F^*(x, \lambda)$ and $F^*(x, \lambda) = F^{f^*}(x, \lambda)$. Therefore, the risk probability policy $\pi^* := \{\hat{f}^*_0, \hat{f}^*_1, \ldots, \hat{f}^*_k, \ldots\}$ is optimal, where $\hat{f}^*_0(x, \lambda) := f^*(x, \lambda)$, $\hat{f}^*_k(h_k) := f^*(x_k, \lambda_k)$ for any $(x, \lambda) \in E \times R, h_k \in H_k, k \geq 1$.

P r o o f. (a) For each $(x, \lambda) \in E \times R, \pi = \{\varphi_0, \varphi_1, \ldots\} \in \Pi$, by using Lemma 3 (b) and (15), we have

$$F^\pi(x, \lambda) = T^{\varphi_0} F^{{}^1\pi}(x, \lambda) \geq T^{\varphi_0} F^*(x, \lambda) \geq TF^*(x, \lambda),$$

which implies that $F^* \geq TF^*$.

On the other hand, by Theorem 2 and (15), we obtain

$$F^*_{n+1}(x,\lambda) = TF^*_n(x,\lambda) \leq T^a F^*_n(x,\lambda). \tag{26}$$

Letting $n \to \infty$ in (26), by using the dominated convergence theorem, we have

$$F^*(x,\lambda) \leq T^a F^*(x,\lambda),$$

which implies $F^* \leq TF^*$. Thus, $F^* = TF^*$ has been proven.

Since $F^* = TF^*$, it follows Lemma 1(b) that there exists a policy $f^* \in F$ such that

$$F^*(x,\lambda) = T^{f^*} F^*(x,\lambda). \tag{27}$$

Suppose that G is another solution to the equation $G(x,\lambda) = TG(x,\lambda)$. Then, based on Lemma 1(b), there exists a policy $f^{'} \in F$ satisfies

$$G(x,\lambda) = T^{f^{'}} G(x,\lambda), \tag{28}$$

which together with (27) and (17) give that

$$\begin{aligned} F^*(x,\lambda) - G(x,\lambda) &\leq \widetilde{T}^{f^{'}}(F^* - G)(x,\lambda), \\ G(x,\lambda) - F^*(x,\lambda) &\leq \widetilde{T}^{f^*}(F^* - U^*)(x,\lambda). \end{aligned}$$

According to Theorem 1, we know that $F^* = G = F^{f^*}$.

(b) For any $(x,\lambda) \in E \times R, h_k \in H_k, k \geq 0$, let $\hat{f}^*_0(x,\lambda) := f^*(x,\lambda), \hat{f}^*_k(h_k) := f^*(x_k,\lambda_k)$, $\pi^* := \{\hat{f}^*_0, \hat{f}^*_1, \ldots, \hat{f}^*_k, \ldots\}$, for all $k \geq 0$. Based on $F^* = F^{f^*}$ and (8), we know that $P^{f^*}_{(x,\lambda)}(\int_0^{+\infty} e^{-\int_0^s \alpha(X_t)\mathrm{d}t} c(X_s, \pi^*_s)\,\mathrm{d}s \neq \int_0^{+\infty} e^{-\int_0^s \alpha(X_t)\mathrm{d}t} c(X_s, f^*)\,\mathrm{d}s) = 0$ and $F^{\pi^*} = F^*$. Therefore, π^* is optimal. □

Based on Theorem 3, we establish the value iteration algorithm for computing the value function and the optimal policy as follows.

The value iteration algorithm

Step 1: (Initialization) For arbitrary small $\varepsilon > 0$ and $(x,\lambda) \in E \times R$, given the initial value $F^*_{-1}(x,\lambda) := I_{[0,\infty)}(\lambda)$.

Step 2: (Iteration) For any $(x,\lambda) \in E \times R, a \in A(x), n \geq -1$, the value F^*_{n+1} is calculated as follows:

$$\begin{aligned} T^a F^*_n(x,\lambda) &= \int_E \int_0^{+\infty} F\Big(y, e^{\alpha(x)u}(\lambda - \frac{c(x,a)(1-e^{-\alpha(x)u})}{\alpha(x)})\Big) \\ &\quad \times Q(\mathrm{d}u, \mathrm{d}y \mid x, a), \\ F^*_{n+1}(x,\lambda) &= \min_{a \in A(x)} \{T^a F^*_n(x,\lambda)\}. \end{aligned}$$

Step 3: (Approximation) If $|F^*_{n+1} - F^*_n| < \varepsilon$, the iterative step stops. Then, the value F^*_{n+1} can be considered as an approximation of the value function F^*. Otherwise, the program returns to step 2 for $n+1$.

4. EXAMPLE

In this section, we illustrate our main results through an example and demonstrate how to use the value iteration algorithm to calculate the value function and the optimal policy.

Example 1. (Optimal control for machine maintenance)

Consider a machine maintenance system with three states: fault (0), normal (1), and good (2). At the initial moment s_0, when the system state is in state $x \in \{1, 2\}$, the decision maker can choose the action a_{x1} to conduct a routine inspection or the action a_{x2} to perform maintenance of the machine. Correspondingly, the system needs to incur certain costs at the rate $c(x, a_{x1})$ or $c(x, a_{x2})$. Subsequently, the system remains in state x until the moment s_1. At the new decision making moment s_1, the system transitions to a new state $y \in \{0, 1, 2\}$ with a certain probability, where the sojourn time in each state follows an arbitrary probability distribution. If the system's state becomes $y = 0$, the decision maker must choose action a_{01} to stop the machine's operation for maintenance and repair. If the system's state enters $y \in \{1, 2\}$, the system will evolve in a similar repetitive manner. We can characterize the state evolution of this system as SMDPs with the state space $E = \{0, 1, 2\}$, the allowable sets of actions $A(x) = \{a_{x1}, a_{x2}\}, x \in \{1, 2\}$ and $A(0) = \{a_{01}\}$, and the discount factor $\alpha(x) = \frac{1}{2^x}$ for any $x \in E$. The transition probabilities are assumed as follows:

$$\begin{aligned}
&p(0|0, a_{01}) = 1, && p(0|1, a_{11}) = \frac{1}{3}, && p(2|1, a_{11}) = \frac{2}{3},\\
&p(0|1, a_{12}) = \frac{1}{2}, && p(2|1, a_{12}) = \frac{1}{2}, && p(0|2, a_{21}) = \frac{3}{5},\\
&p(1|2, a_{21}) = \frac{2}{5}, && p(0|2, a_{22}) = \frac{3}{10}, && p(1|2, a_{22}) = \frac{7}{10}.
\end{aligned} \tag{29}$$

For any $u \in [0, +\infty)$, the semi-Markov decision kernels are given by

$$\begin{aligned}
Q(u, 0|1, a_{11}) &= p(0|1, a_{11})(1 - e^{-0.16u}),\\
Q(u, 2|1, a_{11}) &= p(2|1, a_{11})(1 - e^{-0.16u}),\\
Q(u, 0|1, a_{12}) &= p(0|1, a_{12})(1 - e^{-0.12u}),\\
Q(u, 2|1, a_{12}) &= p(2|1, a_{12})(1 - e^{-0.12u}),\\
Q(u, 0|2, a_{21}) &= p(0|2, a_{21})(1 - e^{-0.03u}),\\
Q(u, 1|2, a_{21}) &= p(1|2, a_{21})(1 - e^{-0.03u}),\\
Q(u, 0|2, a_{22}) &= p(0|2, a_{22})(1 - e^{-0.18u}),\\
Q(u, 1|2, a_{22}) &= p(1|2, a_{22})(1 - e^{-0.18u}),
\end{aligned} \tag{30}$$

and the cost rates are given as follows:

$$c(0, a_{01}) = 0, c(0, a_{02}) = 0, c(1, a_{11}) = 1, c(1, a_{12}) = 2, c(2, a_{21}) = 3, c(2, a_{22}) = 4.$$

The decision maker is primarily concerned with how to select the optimal policy.

Firstly, we need to verify the condition of Lemma 2 to ensure that Assumption 1 holds. For any $0 < \varepsilon < 1$, selecting δ satisfies $\delta < -\frac{25}{9}\ln\varepsilon$, and applying (30), we obtain $Q(\delta, E|x, a) < 1 - \varepsilon$ with $a \in A(x)$. Since the regularity condition of Lemma 2 is satisfied, we can conclude that Assumption 1 holds. Furthermore, based on Theorems 2 and 3, we use the value iteration algorithm to calculate the value function and the optimal policy as follows:

Since $c(0, a_{01}) = 0, 0 \le c(x, a) \le 4$ for any $x \in \{1, 2\}, a \in A(x)$, by (11), it is known that $F^*(0, \lambda) = I_{[0,+\infty)}(\lambda)$ and $F^*(x, \lambda) = 1$ for $\lambda \ge 16, x \in \{1, 2\}$.

Step 1: For $\lambda \in R^+$ and $x \in \{1, 2\}$, let $F^*_{-1}(x, \lambda) := 1$.

Step 2: For $x = 1, n \ge 0$,

$$
\begin{aligned}
T^{a_{11}}F_n^*(1,\lambda) &= \frac{1}{3}\times 0.16\times\int_0^{+\infty} F_n^*(0, e^{0.5u}(\lambda-1+e^{-0.5u}))e^{-0.16u}\mathrm{d}u \\
&\quad +\frac{2}{3}\times 0.16\times\int_0^{+\infty} F_n^*(2, e^{0.5u}(\lambda-1+e^{-0.5u}))e^{-0.16u}\mathrm{d}u \\
&= \begin{cases}
\frac{1}{3}\times(1-e^{0.32\ln(1-\lambda)})+\frac{2}{3}\times 0.16\times\int_0^{-2\ln(1-\lambda)} & \\
\times F_n^*(2, e^{0.5u}(\lambda-1+e^{-0.5u}))e^{-0.16u}\mathrm{d}u, & 0<\lambda<1,\\
\frac{1}{3}+\frac{2}{3}\times F_n^*(2,1), & \lambda=1,\\
\frac{1}{3}+\frac{2}{3}\times(\frac{15}{\lambda-1})^{-0.32}+\frac{2}{3}\times 0.16\times\int_0^{2\ln\frac{15}{\lambda-1}} & \\
F_n^*(2, e^{0.5u}(\lambda-1+e^{-0.5u}))e^{-0.16u}\mathrm{d}u, & 1<\lambda<16,\\
1, & \lambda\ge 16.
\end{cases} \\
&= \begin{cases}
\frac{1}{3}\times(1-e^{0.32\ln(1-\lambda)}) & \\
+\frac{2}{3}\times 0.16\times\int_0^{\lambda} F_n^*(2,t)(\frac{1-t}{1-\lambda})^{-0.32}\frac{2}{1-t}\mathrm{d}t, & 0<\lambda<1,\\
\frac{1}{3}+\frac{2}{3}\times F_n^*(2,1), & \lambda=1,\\
\frac{1}{3}+\frac{2}{3}\times(\frac{15}{\lambda-1})^{-0.32} & \\
+\frac{2}{3}\times 0.16\times\int_{\lambda}^{16} F_n^*(2,t)(\frac{t-1}{\lambda-1})^{-0.32}\mathrm{d}t, & 1<\lambda<16,\\
1, & \lambda\ge 16.
\end{cases}
\end{aligned}
\tag{31}
$$

$$
\begin{aligned}
T^{a_{12}}F_n^*(1,\lambda) &= \frac{1}{2}\times 0.12\times\int_0^{+\infty} F_n^*(0, e^{0.5u}(\lambda-2+2e^{-0.5s}\mathrm{d}s))e^{-0.12u}\mathrm{d}u \\
&\quad +\frac{1}{2}\times 0.12\times\int_0^{+\infty} F_n^*(2, e^{0.5u}(\lambda-2+2e^{-0.5u}))e^{-0.12u}\mathrm{d}u \\
&= \begin{cases}
\frac{1}{2}\times(1-e^{0.24\ln(1-\frac{\lambda}{2})})+\frac{1}{2}\times 0.12\times\int_0^{-2\ln(1-\frac{\lambda}{2})} & \\
F_n^*(2, e^{0.5u}(\lambda-2+2e^{-0.5u}))e^{-0.12u}\mathrm{d}u, & 0<\lambda<2,\\
\frac{1}{2}+\frac{1}{2}\times F_n^*(2,2), & \lambda=2,\\
\frac{1}{2}+\frac{1}{2}\times(\frac{14}{\lambda-2})^{-0.24}+\frac{1}{2}\times 0.12\times\int_0^{2\ln\frac{14}{\lambda-2}} & \\
F_n^*(2, e^{0.5u}(\lambda-2+2e^{-0.5u}))e^{-0.12u}\mathrm{d}u, & 2<\lambda<16,\\
1, & \lambda\ge 16.
\end{cases}
\end{aligned}
$$

$$
\begin{aligned}
&= \begin{cases}
\frac{1}{2} \times (1 - e^{0.24 \ln(1-\frac{\lambda}{2})}) \\
+\frac{1}{2} \times 0.12 \times \int_0^{\lambda} F_n^*(2,t)(\frac{t-2}{\lambda-2})^{-0.24}(\frac{2}{2-t})\mathrm{d}t, & 0 < \lambda < 2, \\
\frac{1}{2} + \frac{1}{2} \times F_n^*(2,2), & \lambda = 2, \\
\frac{1}{2} + \frac{1}{2} \times (\frac{14}{\lambda-2})^{-0.24} \\
+\frac{1}{2} \times 0.12 \times \int_{\lambda}^{16} F_n^*(2,t)(\frac{t-2}{\lambda-2})^{-0.24}(\frac{2}{t-2})\mathrm{d}t, & 2 < \lambda < 16, \\
1, & \lambda \geq 16.
\end{cases} \\
F_{n+1}^*(1,\lambda) &= \min\{T^{a_{11}} F_n^*(1,\lambda), T^{a_{12}} F_n^*(1,\lambda)\}.
\end{aligned}
$$

For $x = 2, n \geq 0$,

$$
\begin{aligned}
T^{a_{21}} F_n^*(2,\lambda) &= \frac{3}{5} \times 0.03 \times \int_0^{+\infty} F_n^*(0, e^{0.25u}(\lambda - 12 + 12e^{-0.25u}))e^{-0.03u}\,\mathrm{d}u \\
&\quad + \frac{2}{5} \times 0.03 \times \int_0^{+\infty} F_n^*(1, e^{0.25u}(\lambda - 12 + 12e^{-0.25u}))e^{-0.03u}\,\mathrm{d}u \\
&= \begin{cases}
\frac{3}{5} \times (1 - e^{0.12 \ln(1-\frac{\lambda}{12})}) \\
+\frac{2}{5} \times 0.03 \times \int_0^{-4\ln(1-\frac{\lambda}{12})} \\
F_n^*(1, e^{0.25u}(\lambda - 12 + 12e^{-0.25u}))e^{-0.03u}\mathrm{d}u, & 0 < \lambda < 12, \\
\frac{3}{5} + \frac{2}{5} \times F_n^*(1,12), & \lambda = 12, \\
\frac{3}{5} + \frac{2}{5} \times (\frac{4}{\lambda-12})^{-0.12} \\
+\frac{2}{5} \times 0.03 \times \int_0^{4\ln\frac{4}{\lambda-12}} \\
F_n^*(1, e^{0.25u}(\lambda - 12 + 12e^{-0.25u}))e^{-0.03u}\mathrm{d}u, & 12 < \lambda < 16, \\
1, & \lambda \geq 16.
\end{cases} \\
&= \begin{cases}
\frac{3}{5} \times (1 - e^{0.12 \ln(1-\frac{\lambda}{12})}) \\
+\frac{2}{5} \times 0.03 \times \int_0^{\lambda} F_n^*(1,t)(\frac{12-t}{12-\lambda})^{-0.12}(\frac{4}{12-t})\mathrm{d}t, & 0 < \lambda < 12, \\
\frac{3}{5} + \frac{2}{5} \times F_n^*(1,12), & \lambda = 12, \\
\frac{3}{5} + \frac{2}{5} \times (\frac{4}{\lambda-12})^{-0.12} \\
+\frac{2}{5} \times 0.03 \times \int_{\lambda}^{16} F_n^*(1,t)(\frac{t-12}{\lambda-12})^{-0.12}(\frac{4}{t-12})\mathrm{d}t, & 12 < \lambda < 16, \\
1, & \lambda \geq 16.
\end{cases}
\end{aligned}
$$

$$
\begin{aligned}
T^{a_{22}} F_n^*(2,\lambda) &= \frac{3}{10} \times 0.18 \times \int_0^{+\infty} F_n^*(0, e^{0.25u}(\lambda - 16 + 16e^{-0.25u}))e^{-0.18u}\mathrm{d}u \\
&\quad + \frac{7}{10} \times 0.18 \times \int_0^{+\infty} F_n^*(1, e^{0.25u}(\lambda - 16 + 16e^{-0.25u}))e^{-0.18u}\mathrm{d}u \\
&= \begin{cases}
\frac{3}{10} \times (1 - e^{0.72 \ln(1-\frac{\lambda}{16})}) \\
+\frac{7}{10} \times 0.18 \times \int_0^{-4\ln(1-\frac{\lambda}{16})} \\
F_n^*(1, e^{0.25u}(\lambda - 16 + 16e^{-0.25u}))e^{-0.18u}\mathrm{d}u, & 0 < \lambda < 16, \\
\frac{3}{10} + \frac{7}{10} \times F_n^*(1,16), & \lambda = 16, \\
1, & \lambda \geq 16.
\end{cases}
\end{aligned}
$$

$$= \begin{cases} \frac{3}{10} \times (1 - e^{0.72\ln(1-\frac{\lambda}{16})}) \\ +\frac{7}{10} \times 0.18 \times \int_0^{\lambda} F_n^*(1,t)(\frac{16-t}{16-\lambda})^{-0.72}(\frac{4}{16-t})\mathrm{d}t, & 0 < \lambda < 16, \\ \frac{3}{10} + \frac{7}{10} \times F_n^*(1,16), & \lambda = 16, \\ 1, & \lambda \geq 16. \end{cases}$$

$$F_{n+1}^*(2,\lambda) = \min\{T^{a_{21}}F_n^*(2,\lambda), T^{a_{22}}F_n^*(2,\lambda)\}.$$

Step 3: Given $\varepsilon = 10^{-12}$, if $|F_{n+1}^* - F_n^*| < \varepsilon$, the program goes to step 4. The value F_{n+1}^* is regarded as an approximation of the value F^*. Otherwise, the program returns to step 2 for $n+1$.

Step 4: Plot the function $T^a F_n^*(x,\lambda)$ and $F^*(x,\lambda)$ for $x \in \{1,2\}$ by using MATLAB software, see Figures 1 and 2.

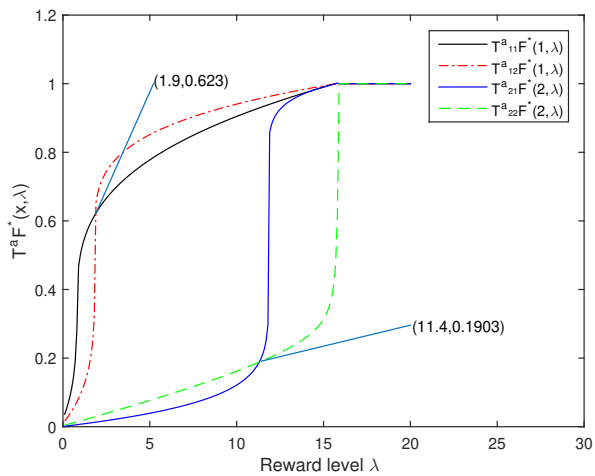

Fig. 1. The function $T^a F^*(x,\lambda)$.

Remark 4. Based on the trapezoidal integration method in [15], we approximate the calculation of the integral in (31) as below:

$$\int_0^{\lambda} F_n^*(2,t)(\frac{1-t}{1-\lambda})^{-0.32}\frac{2}{1-t}\mathrm{d}t$$

$$\approx \sum_{k=0}^{l-1}[F_n^*(2,kh)(\frac{1-kh}{1-\lambda})^{-0.32}\frac{2}{1-kh} + F_n^*(2,(k+1)h)(\frac{1-(k+1)h}{1-\lambda})^{-0.32}\frac{2}{1-(k+1)h}]\frac{h}{2},$$

where $k \leq l, k, l \in \mathbb{N}$, $lh = \lambda$, h denotes the step length, and $\mathbb{N}$ represents the set of all positive integers.

By analyzing the value iteration calculation and Figures 1 – 2, we obtain the following conclusions:

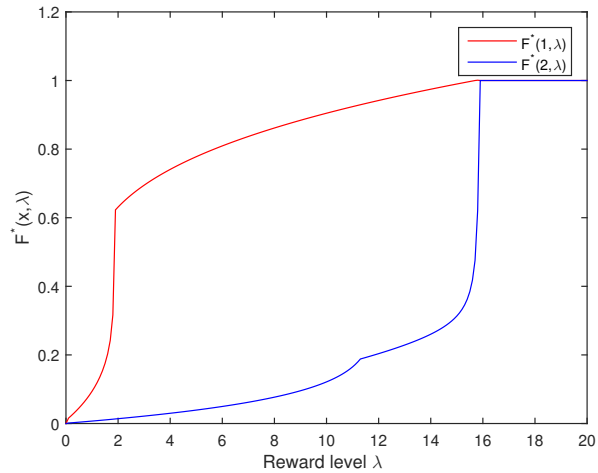

Fig. 2. The value function $F^*(x, \lambda)$.

(a) When the system is in state $x = 1$, if $\lambda \in (0, 1.9)$, the value $T^{a_{12}}F^*(1, \lambda)$ is less than $T^{a_{11}}F^*(1, \lambda)$, it is recommended that the decision maker should choose the low-risk action a_{12}. Otherwise, the decision maker should choose the optimal action a_{11}. When the system state is in state $x = 2$, if $\lambda \in (0, 11.4)$, the decision maker should choose the low-risk action a_{21}. Conversely, if $\lambda \in [11.4, +\infty)$, the decision maker should choose low-risk action a_{22}.

(b) From Figures 1 and 2, it can be seen that when choosing the optimal action, the decision-maker takes into account not only the system state but also the critical points $\lambda^*(1) = 1.9, \lambda^*(2) = 11.4$, where the optimal action is selected as follows:

$$f^*(1, \lambda) = \begin{cases} a_{12}, & 0 \le \lambda < 1.9; \\ a_{11}, & \lambda \ge 1.9. \end{cases}, \quad f^*(2, \lambda) = \begin{cases} a_{21}, & 0 \le \lambda < 11.4; \\ a_{22}, & \lambda \ge 11.4. \end{cases} \tag{32}$$

5. CONCLUSIONS

In this paper, we consider the problem of minimizing the risk probability for infinite discounted Semi-Markov decision processes with varying discount factors. Under the regular condition, we prove that the value function is the unique solution to the optimality equation and the existence of optimal risk probability policy. We provide the value iteration algorithm to approximate the value function and the optimal policy. Our condition is weaker than those in the previous literature. We further investigate the risk probability minimization problem for partially observable semi-Markov decision processes.

ACKNOWLEDGEMENT

This work was supported by Guangxi Natural Science Foundation Program(Grant No.2025GXNSFHA069169); National Natural Science Foundation of China (Grant No.12361091); National Natural Science Foundation of China (Grant No.11961005); Guangxi Natural Science Foundation Program(Grant No.2020GXNSFAA297196); Guangxi science and technology base and talent project(Grant No.AD21159005); The Doctoral Foundation of Guangxi University of Science and Technology (Grant No.18Z06).

(Received December 30, 2024)

REFERENCES

[1] D. Bertsekas and S. E. Shreve: Stochastic Optimal Control: The Discrete-Time Case. Academic Press Inc, New York 1996.

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

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

[4] 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.1100.04779

[5] X. P. Guo and O. Hernández-Lerma: Continuous-time markov decision processes: theory and applications. Springer-Verlag, Berlin 2009.

[6] X. P. Guo, X. Y. Song, and Y. Zhang: First passage optimality for continuous-time Markov decision processes with varying discount factors and history-dependent policies. IEEE Trans. Automat. Control *59* (2013), 163–174. DOI:10.1109/TAC.2013.2281475

[7] O. Hernández-Lerma and J. B. Lasserre: Discrete-time Markov control processes: basic optimality criteria. Springer-Verlag, New York 1996.

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

[9] Y. H. Huang and X. P. Guo: Finite horizon semi-Markov decision processes with application to maintenance systems. European J. Oper. Res. *212* (2011), 131–140. DOI:10.1016/j.ejor.2011.01.027

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

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

[12] X. X. Huang, X. L. Zuo, 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

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

[14] H. F. Huo and X. P. Guo: Risk probability minimization problems for continuous-time Markov decision processes on finite horizon. IEEE Trans. Autom. Control *65* (2019), 3199–3206. DOI:10.1109/TAC.2019.2947654

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

[16] Y. L. Lin, R. J. Tomkins, and C. L. Wang: Optimal models for the first arrival time distribution function in continuous time With a special case. Acta. Math. Appl. Sinica *10* (1994), 194–212.

[17] J. W. Mamer: Successive approximations for finite horizon, semi-Markov decision processes with application to asset liquidation. Oper. Res. *34* (1986), 638–644. DOI:10.1287/opre.34.4.638

[18] M. Sakaguchi and Y. Ohtsubo: Optimal threshold probability and expectation in semi-Markov decision processes. Appl. Math. Comput. *216* (2010), 2947–2958. DOI:10.1016/j.amc.2010.04.007

[19] M. J. Sobel: The variance of discounted Markov decision processes. J. Appl. Probab. *19* (1982), 794–802. DOI:10.1017/s0021900200023123

[20] V. Nollau: Solution of a discounted semi-markovian descision problem by successive oevarrelaxation. Optimization *39* (1997), 85–97. DOI:10.1080/02331939708844273

[21] Y. Ohtsubo: Optimal threshold probability in undiscounted Markov decision processes with a target set. Appl. Math. Anal. Comp. *149* (2004), 519–532. DOI:10.1016/S0096-3003(03)00158-9

[22] A. Piunovskiy, Y. Zhang, and A. N. Shiryaev: Continuous-Time Markov Decision Processes: Borel Space Models and General Control Strategies. Springer, Berlin 2020. DOI:10.1007/978-3-030-54987-9

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

[24] X. Wen, H. F. Huo, and X.P. Guo: First passage risk probability minimization for piecewise deterministic Markov decision processes. Acta Math.Appl.Sin.Engl.Ser. *38* (2022), 549–567. DOI:10.1007/s10255-022-1098-0

[25] C. Wu and Y. 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

[26] X. Wu and X. P. Guo: First passage optimality and variance minimisation of Markov decision processes with varying discount factors. J. Appl. Prob. *52* (2015), 441–456. DOI:10.1239/jap/1437658608

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

Jinhua Cui, School of Science, Guangxi University of Science and Technology, Liuzhou, 545006. P. R. China.
e-mail: liveinsidecoco@163.com

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