Optimal Stopping Time on Semi-Markov Processes with Finite Horizon

Fang Chen¹ · Xianping Guo¹ · Zhong-Wei Liao²

Received: 31 August 2021 / Accepted: 20 March 2022 / Published online: 21 April 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
In this paper, we consider the optimal stopping problems on semi-Markov processes (sMPs) with finite horizon and aim to establish the existence and algorithm of optimal stopping times. The key method is the equivalence between optimal stopping problems on sMPs and a special class of semi-Markov decision processes (sMDPs). We first introduce the optimality equation and show the existence of the optimal policies of finite-horizon sMDPs with additional terminal costs. Based on the optimal stopping problems on sMPs, we give an explicit construction of sMDPs such that the optimal stopping times of sMPs are equivalent to the optimal policies of the constructed sMDPs. Then, using the results of sMDPs developed here, we not only prove the existence of the optimal stopping times of sMPs, but also provide an algorithm for computing the optimal stopping times of sMPs. Moreover, we show that the optimal and ε-optimal stopping time can be characterized by the hitting time of some special sets. Finally, we give an example to illustrate the effectiveness of our conclusions.

Keywords Optimal stopping time · Semi-Markov processes · Semi-Markov decision processes · Optimal policy · Optimality equation · Iterative algorithm

Mathematics Subject Classification 60G40 · 60K15 · 90C40

Communicated by Jörg Rambau.

Zhong-Wei Liao
zhwliao@hotmail.com, zhwliao@bnu.edu.cn
Fang Chen
chenf76@mail2.sysu.edu.cn
Xianping Guo
mcsxgp@mail.sysu.edu.cn

¹ School of Mathematics, Sun Yat-Sen University, Guangzhou 510275, China
² Beijing Normal University, Zhuhai 519087, China

Springer
1 Introduction

The optimal stopping problem is an important branch of the intersection of probability and control theory, which aims to find the optimal stopping time of stochastic systems under some certain criterion, and has been widely studied and applied in finance, such as the pricing of American options and the buying-selling problem; see the monographs \([3, 6, 9, 11, 30]\) and the references therein. In most existing literature on optimal stopping problems, the case on discrete-time Markov processes (Dufour & Piuovskiy [12]; Huang & Zhou [18]; Nute & Zhang [25]; de Saporta et al. [28]), the case on continuous-time Markov processes (Arkin & Slastnikov [1]; Bäuerle & Popp [2]; Christensen & Lindensjö [10]; Gapeev et al. [14]; Shao & Tian [29]), and the case on renewal processes (Karpowicz [20]; Karpowicz & Szajowski [21]; Szajowski [32]) are commonly considered.

Note that the sojourn times between two jump epochs in discrete-time Markov processes, continuous-time Markov processes, and renewal processes are constant, exponentially distributed, and independent and identically distributed, respectively \([13, 20, 32]\). As is well known, semi-Markov processes (sMPs) are more general than these processes since the time can not only follow more distribution but also depend on the present state and the next state. Thus, sMPs have been widely studied and applied in many areas \([17, 19, 24]\). To the best of our knowledge, the optimal stopping problem on sMPs is addressed only in Boshuizen & Gouweleeuw [5], Chen et al. [8], Ohtsubo [26], and Zuckerman [33]. More precisely, Zuckerman [33] studied optimal stopping problem and gave the optimal stopping times with time-dependent costs and the assumption that sojourn times are independent of the next state; Boshuizen & Gouweleeuw [5] extended the results in [33] to the general case where sojourn times depend on the next state; Chen et al. [8] and Ohtsubo [26] both studied the optimal stopping problems with infinite-horizon and state-dependent costs and showed the existence of an optimal time under discounted [8] and undiscounted criterion [26], respectively. However, the underlying process may be forced to end at a certain time in practice. Thus, it is natural and desirable to consider the optimal stopping problem on sMPs with finite horizon and state-dependent costs, which has not been studied yet and will be taken into consideration in this paper.

In many practical applications, if someone violates the contract, he/she must pay additional penalties, which will be regarded as terminal costs. Thus, in our model, the costs consist of both the running costs and additional terminal costs \([3]\). To deal with the optimal stopping problem on sMPs with finite horizon, we follow the idea of transforming the discrete-time optimal stopping problem into an equivalent discrete-time Markov decision process in [3], which has an obvious advantage that the existence and computation of optimal stopping times can be obtained together. Indeed, noting that “continue” or “stop” may be considered as an action, using the date of the optimal stopping problems on sMPs, we also construct corresponding sMDPs with the action space \([0, 1]\), where the action 0 and action 1 mean continuation and stop, respectively. Then, we prove the equivalence between the optimal stopping time problems on sMPs and the corresponding sMDPs, which means that for any stopping time and planning horizon, it can be induced a policy with the same finite-horizon expected cost as that of the stopping time. Furthermore, to deal with the optimal stopping problems on
sMPs with terminal costs by the equivalent sMDPs, we need to extend the results in [17] without any terminal cost to the more natural case with additional terminal costs and establish the existence of an optimal policy and an approximation algorithm for the value function of the sMDPs by the minimal nonnegative solution method. Using this equivalence and the results about sMDPs developed here, we not only show the existence of optimal stopping times on sMPs, but also provide an algorithm for computing optimal stopping times on sMPs, whereas most literature only considers the existence of optimal stopping times. Finally, we prove that the optimal and \( \varepsilon \)-optimal stopping time can be characterized by the hitting time of some special sets and the number of iterations can be estimated explicitly. At the end of the paper, we illustrate the effectiveness of our results through an example of house rental problem.

The rest of the paper is organized as follows. We describe optimal stopping problems on sMPs with finite horizon in Sect. 2. In Sect. 3, we develop some results on sMDPs with additional terminal costs. Our main results on the existence and computation of optimal stopping times are given in Sect. 4 after giving the preliminaries in Sect. 3. In Sect. 5, we show how to use the iterative algorithm to compute the \( \varepsilon-T \)-optimal stopping time through a house rental problem. At the end of this paper, we give a brief conclusion and the prospect of further research work.

2 Optimal Stopping Problem on Semi-Markov Processes

Notation. The following notations are frequently used throughout this paper. \( \mathbb{I}_A \) stands for the indicator function of some set \( A \). \( \delta_x(\cdot) \) is the Dirac measure concentrated at \( x \). \( B(X) \) is the Borel \( \sigma \)-algebra of the Borel space \( X \), and \( \mathcal{P}(X) \) is the set of all probability measures on \( B(X) \). \( \mathbb{R} := (-\infty, +\infty) \), \( \mathbb{R}_+ := [0, +\infty) \). \( x \wedge y := \min\{x, y\} \), \( x^+ := \max\{x, 0\} \). We always use the convention \( \sum_{k=n}^m a_k = 0 \) and \( \prod_{k=n}^m a_k = 1 \) if \( n > m \).

The model of sMPs is \( \{E, Q(\cdot, \cdot|x)\} \), where \( E \) is the state space, which is assumed to be a Borel space equipped with the Borel \( \sigma \)-algebra \( B(E) \). The transition mechanism of the sMPs is defined by the semi-Markov kernel \( Q(\cdot, \cdot|x) \) on \( \mathbb{R}_+ \times E \) given \( E \), which is assumed that:

(i) given any \( B \in \mathcal{B}(E) \) and \( x \in E \), \( Q(\cdot, B|x) \) is a non-decreasing right continuous real-valued function on \( \mathbb{R}_+ \) and satisfies \( Q(0, B|x) = 0 \);
(ii) given any \( t \in \mathbb{R}_+ \), \( Q(t, \cdot|\cdot) \) is a sub-stochastic kernel on \( E \);
(iii) \( \lim_{t \to \infty} Q(t, \cdot|\cdot) \) is a stochastic kernel on \( E \).

We introduce the measurable space \( (\Omega, \mathcal{F}) \), which is based on the Kitaev construction (see [22, 23]). Let

\[
\Omega := \{(x_0, t_1, x_1, \ldots, t_n, x_n, \ldots) : x_0 \in E, t_n \in \mathbb{R}_+, x_n \in E, n \geq 1\},
\]

and \( \mathcal{F} \) be the corresponding product Borel \( \sigma \)-algebra. The history of sMPs up to the \( n \)th jump epoch is

\[
h_0 = x_0, \quad h_{n+1} = (x_0, t_1, x_1, \ldots, t_{n+1}, x_{n+1}), \quad n \geq 0.
\]
Let $H_n$ be the set of all histories $h_n$. For each $\omega = (x_0, t_1, \ldots, x_n, t_{n+1}, \ldots) \in \Omega$, define

$$X_n(\omega) = x_n, \quad T_0(\omega) = 0, \quad T_{n+1}(\omega) = t_{n+1}, \quad S_n(\omega) = \sum_{k=0}^{n} T_k(\omega), \quad \forall n \geq 0,$$

where $S_n$, $T_{n+1}$, and $X_n$ are the $n$th jump time, the sojourn time between the $n$th and $(n+1)$th jumps, and the state at the $n$th jump time, respectively. Further, we assume that the decision may only depend on the observation of the marked point process $\{T_n, X_n, n \geq 0\}$. Denote by $\mathcal{F}_n$, $n \geq 0$ the filtration generated by $\{T_n, X_n, n \geq 0\}$, i.e.,

$$\mathcal{F}_n := \sigma(T_0, X_0, \ldots, T_n, X_n).$$

We give the definition of stopping times with respect to the filtration $\{\mathcal{F}_n, n \geq 0\}$ as the following.

**Definition 2.1** A random variable $\tau : \Omega \to \mathbb{N} \cup \{+\infty\}$ is called $\mathcal{F}_n$-stopping time if $\{\tau = n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$. Denote by $\Gamma$ the set of all $\mathcal{F}_n$-stopping times.

This definition means that upon observing the marked point process $\{X_n, T_n, n \geq 0\}$ until $n$th jump time, we can decide whether $\tau$ has already occurred. In the absence of ambiguity, we write $\mathcal{F}_n$-stopping time as stopping time. Using the Tulcea theorem (see [16, Proposition C.10]), for each $x \in E$, there exists a unique probability measure $P_x$ on $(\Omega, \mathcal{F})$ satisfying that $P_x(T_0 = 0, X_0 = x) = 1$ and

$$P_x(T_{n+1} \leq t, X_{n+1} \in B|Y_n) = Q(t, B|X_n),$$

where $Y_n := (X_0, T_1, X_1, \ldots, T_n, X_n)$. Denote by $E_x$ the expectation with respect to $P_x$. Moreover, we give the following standard condition, which ensures the regularity of sMPs and is widely used in sMPs and sMDPs; see [8, 17, 24] for instance.

**Assumption 2.2** There exist $\delta > 0$ and $\alpha > 0$, such that

$$Q(\delta, E|x) \leq 1 - \alpha \quad \forall x \in E. \quad (2.1)$$

According to [17], Assumption 2.2 implies the regularity of sMPs, i.e.,

$$P_x( \lim_{n \to \infty} S_n = \infty ) = 1 \quad \forall x \in E.$$

Corresponding to $\{(T_n, X_n), n \geq 0\}$, we define an underlying continuous-time process $\{X(t), t \in \mathbb{R}_+\}$ by

$$X(t) = X_n, \quad S_n \leq t < S_{n+1}. \quad (2.2)$$

Refer to Limnios and Oprisan [24] for more details about $\{X(t), t \in \mathbb{R}_+\}$. In the next step, we introduce the optimal stopping problem with finite horizon.
Fix any planning horizon $T \in \mathbb{R}_+$. For each stopping time $\tau \in \Gamma$, the total cost is given as

\[
R^T_\tau := \begin{cases} 
\int_0^{S_\tau} c(X(t))dt + g(X(S_\tau)), & S_\tau < T; \\
\int_0^T c(X(t))dt, & S_\tau \geq T.
\end{cases}
\] (2.3)

where the functions $c(\cdot)$ and $g(\cdot)$ are the cost rates and the terminal costs, respectively, which are nonnegative real-valued measurable functions on $E$. This total cost (2.3) means that we either stop after time $T$ and pay the accumulated costs on $[0, T]$, or stop before time $T$ and pay two parts of the cost, which include the accumulated costs on $[0, S_\tau]$ and the terminal cost $g(X(S_\tau))$. This is very common in practical applications, such as house rental problems. Because if someone violates the contract, he/she must pay an additional penalty, which is measured by the terminal cost.

For each stopping time $\tau \in \Gamma$, the $T$-horizon expected cost is defined as

\[
V^\tau(T, x) := \mathbb{E}_x[R^T_\tau], \quad x \in E.
\] (2.4)

Then, the value function of the optimal stopping problems with finite horizon $T$ is

\[
V^*(T, x) := \inf_{\tau \in \Gamma} V^\tau(T, x).
\] (2.5)

**Definition 2.3** A stopping time $\tau^* \in \Gamma$ is called $T$-optimal if it satisfies that

\[
V^{T^*}(T, x) = V^*(T, x) = \inf_{\tau \in \Gamma} V^\tau(T, x) \quad \forall x \in E.
\]

The main purpose of this paper is to find the $T$-optimal stopping times and give an algorithm for computing the value function $V^*$. The key method is the observation that finding the $T$-optimal stopping times on sMPs is equivalent to finding the $T$-optimal policies, see Definition 3.6, in a special class of semi-Markov decision processes (sMDPs). In the following section, we introduce the main results of the finite-horizon sMDPs with additional terminal costs, including the optimality equation and the existence of the $T$-optimal policies; see Theorem 3.11.

### 3 Finite-Horizon Semi-Markov Decision Processes with Terminal Costs

We will establish the relationship between optimal stopping problems on sMPs and sMDPs. To this end, we first introduce the model of sMDPs with terminal costs and give the optimality equation and the $T$-optimal policies of this model. In particular, if the terminal costs are equal to 0, our results are degenerated into the classical case in [17]. Here and in what follows, we use “^*” to distinguish sMDPs from sMPs.
The model of sMDPs is introduced as
\[ \{ \hat{E}, A, (A(x), x \in \hat{E}), \hat{Q}(\cdot, \cdot|x, a), \hat{c}(x, a), \hat{g}(x, a) \}, \]
where \( \hat{E} \) is the state space and \( A \) is the action set, which are assumed to be a Borel space and a denumerable set, respectively. \( A(x) \) is the set of all admissible actions at the state \( x \in \hat{E} \), which is assumed to be finite. Let \( K := \{ (x, a) | x \in \hat{E}, a \in A(x) \} \) be the set of all admissible state-action pairs, and the function \( \hat{Q}(\cdot, \cdot|x, a) \) is the semi-Markov kernel on \( \mathbb{R}_+ \times \hat{E} \) given \( K \). Assume that \( K \in \mathcal{B}(\hat{E}) \times \mathcal{B}(A) \) and there is a measurable mapping \( f : \mathbb{R} \times \hat{E} \to A \) such that \( (x, f(t, x)) \in K \) for all \( (t, x) \in \mathbb{R} \times \hat{E} \). Finally, the functions \( \hat{c}(x, a) \) and \( \hat{g}(x, a) \) on \( K \) represent the cost rates and terminal costs, which are assumed to be nonnegative and measurable.

**Remark 3.1** In particular, when \( \hat{E} \) is denumerable and \( \hat{g} \equiv 0 \), the model given above degenerates to the case studied in [17].

The evolution of the finite-horizon sMDPs is as follows. Initially, the system occupies some state \( x_0 \in \hat{E} \) and the decision-maker has a planning horizon \( s \in \mathbb{R} \); then, he/she chooses an action \( a_0 \in A(x_0) \) according to the state \( x_0 \) and the planning horizon \( s \). As a consequence of this action choice, the system jumps to a state \( x_1 \) after the sojourn time \( t_1 \), in which the transition law is subject to the semi-Markov kernel \( \hat{Q} \). At time \( t_1 \), there is a remaining planning horizon \( s - t_1 \) for the decision-maker. According to the current state \( x_1 \), the current planning horizon \( s - t_1 \) as well as the previous state and action \( (x_0, a_0) \), the decision-maker chooses an action \( a_1 \in A(x_1) \) and the same sequence of events occurs. The system of sMDPs evolves in this way, and then at the \( n \)th decision epoch, we obtain a remaining planning horizon \( s - \sum_{k=1}^{n} t_k \) and an admissible history \( \hat{h}_n \) with the form
\[
\hat{h}_n = (x_0, a_0, t_1, x_1, \ldots, a_{n-1}, t_n, x_n),
\]
where \( (x_m, a_m) \in K, t_{m+1} \in \mathbb{R}_+ \) for all \( m = 0, 1, \ldots, n - 1 \), \( x_n \in \hat{E} \). Let \( \hat{H}_n \) denote the set of all admissible histories \( \hat{h}_n \) of the system up to the \( n \)th decision epoch, which is endowed with the Borel \( \sigma \)-algebra.

**Definition 3.2** A policy \( \pi = \{ \pi_n, n \geq 0 \} \) is a sequence of stochastic kernels \( \pi_n \) on \( A \) given \( \mathbb{R} \times \hat{H}_n \) satisfying
\[
\pi_n(A(x_n)|s, \hat{h}_n) = 1, \quad \forall n \geq 0, \quad s \in \mathbb{R}, \quad \hat{h}_n = (x_0, a_0, t_1, x_1, \ldots, a_{n-1}, t_n, x_n) \in \hat{H}_n.
\]
(3.1)

Denote by \( \Pi \) the set of all policies.

**Remark 3.3** The parameter \( s \) in (3.1) represents the remaining planning horizon up to the \( n \)th decision epoch. For convenience, this parameter is allowed to be negative. By Definition 3.2, the policies in our model depend on the remaining planning horizon; as a contrast, the policies are independent of the remaining horizon in the infinite-horizon model; see [7, 31].
To distinguish the subclasses of $\Pi$, we introduce the following notations. Let $\Phi$ represent the set of stochastic kernels $\varphi$ on $A$ given $\mathbb{R} \times \hat{E}$ such that $\varphi(A(x)|s, x) = 1$ for all $(s, x) \in \mathbb{R} \times \hat{E}$, and $\mathbb{F}$ represent the set of measurable functions $f : \mathbb{R} \times \hat{E} \rightarrow A$ such that $f(s, x) \in A(x)$ for all $(s, x) \in \mathbb{R} \times \hat{E}$.

**Definition 3.4** (a) A policy $\pi = \{\pi_n\}$ is said to be Markov if there is a sequence $\{\varphi_n\}$ of stochastic kernels $\varphi_n \in \Phi$ such that $\pi_n(\cdot|s, \hat{H}_n) = \varphi_n(\cdot|s, x_n)$ for every $(s, \hat{H}_n) \in \mathbb{R} \times \hat{H}_n$ and $n \geq 0$. We write such a policy as $\pi = \{\varphi_n\}$.

(b) A Markov policy $\pi = \{\varphi_n\}$ is said to be stationary if $\varphi_n$ are independent of $n$. In this case, we write $\pi$ as $\varphi$ for simplicity.

(c) A policy $\pi = \{\pi_n\}$ is called deterministic if there exists a sequence $\{d_n\}$ of measurable functions $d_n : \mathbb{R} \times \hat{H}_n \rightarrow A$ such that for all $(s, \hat{H}_n) \in \mathbb{R} \times \hat{H}_n$, $d_n(s, \hat{H}_n) \in A(x_n)$

$$\pi_n(a|s, \hat{H}_n) = \delta_{\{d_n(s, \hat{H}_n)\}}(a) \quad \forall a \in A.$$ We write such a policy as $\pi = \{d_n\}$.

(d) A Markov policy $\pi = \{\varphi_n\}$ is said to be deterministic Markov if there is a sequence $\{f_n\}$ of functions $f_n \in \mathbb{F}$ such that $\varphi_n(\cdot|s, x)$ is concentrated at $f_n(s, x)$ for all $(s, x) \in \mathbb{R} \times \hat{E}$ and $n \geq 0$. We write such a policy as $\pi = \{f_n\}$.

(e) A deterministic Markov policy $\pi = \{f_n\}$ is said to be deterministic stationary if $f_n$ are independent of $n$. In this case, we write $\pi$ as $f$ for simplicity.

For convenience, we denote by $\Pi_{RM}$, $\Pi_{RS}$, $\Pi_{DH}$, $\Pi_{DM}$, and $\Pi_{DS}$ the sets of all Markov, stationary, deterministic, deterministic Markov, and deterministic stationary policies, respectively. We can verify that $\mathbb{F} = \Pi_{DS} \subset \Pi_{DM} \subset \Pi_{DH} \subset \Pi$ and $\mathbb{F} \subset \Phi = \Pi_{RS} \subset \Pi_{RM} \subset \Pi$.

Let $\hat{\Omega} = (\hat{E} \times A \times \mathbb{R}_+)^{\infty}$ be a sample space and $\hat{\mathcal{F}}$ be the corresponding product $\sigma$-algebra. Similar to sMPS, for any $\hat{\omega} = (x_0, a_0, t_1, \ldots, x_n, a_n, t_{n+1}, \ldots) \in \hat{\Omega}$ and $n \geq 0$, we define

$$\hat{T}_0(\hat{\omega}) = 0, \quad \hat{T}_n(\hat{\omega}) = t_n, \quad \hat{X}_n(\hat{\omega}) = x_n, \quad A_n(\hat{\omega}) = a_n.$$ For each $n \geq 0$, let $\hat{S}_n = \sum_{m=0}^{n} \hat{T}_m$, and then we define $\{\hat{X}(t), A(t), t \in \mathbb{R}_+\}$ by

$$\hat{X}(t) := \begin{cases} \hat{X}_n, & \hat{S}_n \leq t < \hat{S}_{n+1}; \\ \partial_S, & t \geq \lim_{n \to \infty} \hat{S}_n, \end{cases} \quad A(t) := \begin{cases} A_n, & \hat{S}_n \leq t < \hat{S}_{n+1}; \\ \partial_A, & t \geq \lim_{n \to \infty} \hat{S}_n, \end{cases}$$

where $\partial_S$ and $\partial_A$ are the extra state and action jointed to $\hat{E}$ and $A$, respectively. Now, given $(s, x) \in \mathbb{R} \times \hat{E}$ and $\pi = \{\pi_n\} \in \Pi$, by the Tulcea Theorem (see [16, Proposition C.10]), there exists a unique probability measure $\hat{P}_{(s,x)}^\pi$ on $(\hat{\Omega}, \hat{\mathcal{F}})$ such that

$$\hat{P}_{(s,x)}^\pi(\hat{T}_0 = 0, \hat{X}_0 = x) = 1, \quad \hat{P}_{(s,x)}^\pi(A_n = a|\hat{Y}_n) = \pi_n(a|s - \hat{S}_n, \hat{Y}_n), \quad a \in A,$$
\begin{align}
\hat{\mathbb{P}}^{\pi}_{(s,x)}(\hat{T}_{n+1} \leq t, \hat{X}_{n+1} \in B | \hat{Y}_n, A_n) &= \hat{Q}(t, B | \hat{X}_n, A_n), \quad t \in \mathbb{R}_+, \ B \in \mathcal{B}(\hat{E}), \ a \in A.
\end{align}

where $\hat{Y}_n := (\hat{X}_0, A_0, \hat{T}_1, \hat{X}_1, \ldots, A_{n-1}, \hat{T}_n, \hat{X}_n)$. The expectation with respect to $\hat{\mathbb{P}}^{\pi}_{(s,x)}$ is denoted by $\hat{\mathbb{E}}^{\pi}_{(s,x)}$.

The following assumption is a regularity condition of the controlled system. Intuitively, we need to eliminate the explosion case of $T$-horizon sMDPs, that is, the condition is needed to ensure that the system will not jump infinitely in a finite time, and the decision-maker will not make infinite decisions in a finite time. Thus, we propose the following regularity assumption.

**Assumption 3.5** For all $(s, x) \in \mathbb{R} \times \hat{E}$ and $\pi \in \Pi$, $\hat{\mathbb{P}}^{\pi}_{(s,x)} \left( \left\{ \lim_{n \to \infty} \hat{S}_n = \infty \right\} \right) = 1$.

Given $(s, x) \in \mathbb{R}_+ \times \hat{E}$, the $s$-horizon expected cost of a policy $\pi \in \Pi$ is defined as

\[ U^\pi(s, x) := \hat{\mathbb{E}}^{\pi}_{(s,x)} \left[ \int_0^s \hat{c}(\hat{X}(t), A(t))dt + \hat{g}(\hat{X}(s), A(s)) \right], \]

and the value function (or minimum expected cost) is $U^*(s, x) := \inf_{\pi \in \Pi} U^\pi(s, x)$.

**Definition 3.6** A policy $\pi^* \in \Pi$ is called $T$-optimal if

\[ U^{\pi^*}(s, x) = U^*(s, x) \quad \forall (s, x) \in [0, T] \times \hat{E}. \]

**Remark 3.7** Note that the stopping time $\tau^*$ in sMPs is $T$-optimal if and only if it achieves $V^*$ for the fixed planning horizon $T$. As a comparison, the policy $\pi^* \in \Pi$ in sMDPs is $T$-optimal if and only if it achieves $U^*$ for all $s \in [0, T]$.

The focus of this section is finding a $T$-optimal policy in $\Pi$, which can induce the $T$-optimal stopping time in sMPs. The following result implies that it suffices to seek for the $T$-optimal policy in the subset $\Pi_{RM}$ instead of $\Pi$.

**Proposition 3.8** Suppose that Assumption 3.5 holds. For each $\pi = \{\pi_n\} \in \Pi$ and $(s, x) \in \mathbb{R}_+ \times \hat{E}$, there exists a policy $\hat{\pi} = \{\phi_n\} \in \Pi_{RM}$ such that $U^\pi(s, x) = U^{\hat{\pi}}(s, x)$.

**Proof** Under Assumption 3.5, the monotone convergence theorem gives that

\[ U^\pi(s, x) = \sum_{m=0}^{\infty} \hat{\mathbb{E}}^{\pi}_{(s,x)} \left[ (s - \hat{S}_m)^+ \wedge \hat{T}_{m+1} \right] \hat{c}(\hat{X}_m, A_m) \]

\[ + \mathbf{1}_{[\hat{S}_m, \hat{S}_{m+1})}(s) \hat{g}(\hat{X}_m, A_m) \]  

Hence, it suffices to show that for each $\pi \in \Pi$, there is a policy $\hat{\pi} = \{\phi_n\} \in \Pi_{RM}$ such that

\[ \hat{\mathbb{P}}^{\pi}_{(s,x)}(\hat{X}_n \in B, A_n = a, \hat{S}_n \leq t, \hat{T}_{n+1} \leq v) \]
\[
\hat{\pi}_{(s,x)}(\hat{X}_n \in B, A_n = a, \hat{S}_n \leq t, \hat{T}_{n+1} \leq v), \quad (3.6)
\]

for all \(n \in \mathbb{N}, t, v \in \mathbb{R}_+, B \in \mathcal{B}(\hat{E})\) and \(a \in A\). Note that (3.4) implies

\[
\hat{E}_\pi^{\pi}(s,x) \left[ \mathbb{1}_{[0,v]}(\hat{T}_{n+1}|\hat{X}_n, A_n, \hat{S}_n) \right] = \hat{Q}(v, \hat{E}|\hat{X}_n, A_n).
\]

Hence, in order to prove (3.6), we only need to verify the following equation:

\[
\hat{E}_\pi^{\pi}(s,x) (\hat{X}_n \in B, \hat{S}_n \leq t, A_n = a) = \hat{E}_\pi^{\pi}(s,x) (\hat{X}_n \in B, \hat{S}_n \leq t, A_n = a). \quad (3.7)
\]

Indeed, for each \((s, x) \in \mathbb{R}_+ \times \hat{E}\), we define a Markov policy \(\hat{\pi} := (\varphi_n)\) as

\[
\varphi_n(a|t, y) := \begin{cases} 
\hat{E}_\pi^{\pi}(s,x) \left[ \mathbb{1}_{[a]}(A_n)|\hat{S}_n = s - t, \hat{X}_n = y \right], & t \leq s, y \in \hat{E}; \\
|A(y)|^{-1}, & t > s, y \in \hat{E};
\end{cases} \quad (3.8)
\]

where \(|A(y)|\) is the cardinality of \(A(y)\). Next, we show (3.7) holds with \(\hat{\pi}\) defined in (3.8) by induction. It is easy to verify that (3.7) holds with \(n = 0\). Assume that (3.7) holds for some \(n (n \geq 0)\). Then,

\[
\hat{E}_\pi^{\pi}(s,x) (\hat{X}_{n+1} \in B, \hat{S}_{n+1} \leq t) = \hat{E}_\pi^{\pi}(s,x) \left[ \hat{E}_\pi^{\pi}(s,x) \left[ \mathbb{1}_B(\hat{X}_{n+1}) \mathbb{1}_{[0,t]}(\hat{S}_{n+1}) | \hat{X}_n, A_n, \hat{S}_n \right] \right] \\
= \hat{E}_\pi^{\pi}(s,x) \left[ \hat{Q}(t - \hat{S}_n)^+, B|\hat{X}_n, A_n \right] \quad \text{(by (3.4))} \\
= \hat{E}_\pi^{\pi}(s,x) (\hat{X}_{n+1} \in B, \hat{S}_{n+1} \leq t). \quad \text{(by the induction hypothesis)} \quad (3.9)
\]

Therefore, the definition of \(\hat{\pi}\) and the above equality imply that

\[
\hat{E}_\pi^{\pi}(s,x) \left( \hat{X}_{n+1} \in B, \hat{S}_{n+1} \leq t, A_{n+1} = a \right) \nonumber \\
= \hat{E}_\pi^{\pi}(s,x) \left[ \mathbb{1}_B(\hat{X}_{n+1}) \mathbb{1}_{[0,t]}(\hat{S}_{n+1}) \hat{E}_\pi^{\pi}(s,x) \left[ \mathbb{1}_{[a]}(A_{n+1}) | \hat{S}_{n+1}, \hat{X}_{n+1} \right] \right] \\
= \hat{E}_\pi^{\pi}(s,x) \left[ \mathbb{1}_B(\hat{X}_{n+1}) \mathbb{1}_{[0,t]}(\hat{S}_{n+1}) \varphi_n(a|s - \hat{S}_{n+1}, \hat{X}_{n+1}) \right] \quad \text{(by (3.8))} \\
= \hat{E}_\pi^{\pi}(s,x) \left[ \mathbb{1}_B(\hat{X}_{n+1}) \mathbb{1}_{[0,t]}(\hat{S}_{n+1}) \hat{E}_\pi^{\pi}(s,x) \left[ \mathbb{1}_{[a]}(A_{n+1}) | \hat{S}_{n+1}, \hat{X}_{n+1} \right] \right] \quad \text{(by (3.3))} \\
= \hat{E}_\pi^{\pi}(s,x) \left( \hat{X}_{n+1} \in B, \hat{S}_{n+1} \leq t, A_{n+1} = a \right). \quad \text{(by (3.9))}
\]

By induction, equation (3.7) holds for all \(n \geq 0\) and the proof is completed. \(\square\)

Due to Proposition 3.8, we only need to consider Markov policies in the rest of this section. Next, we introduce the main results on the finite-horizon sMDPs. In detail, we prove that the value function is the minimum nonnegative solution to the optimality equation and there exists a \(T\)-optimal deterministic stationary policy. Furthermore, we derive an algorithm for computing the \(T\)-optimal policies and the value function.
Let $\mathbb{M}$ be the set of Borel measurable functions $v : [0, T] \times \hat{E} \rightarrow [0, \infty]$. Given any $a \in A$, define an operator $\mathbb{T}^a : \mathbb{M} \rightarrow \mathbb{M}$ as: for each $v \in \mathbb{M}$ and $x \in \hat{E}$, $\mathbb{T}^a v(s, x) := +\infty$ when $a \not\in A(x)$, otherwise

\[
\mathbb{T}^a v(s, x) := \hat{c}(x, a) \int_0^s \left( 1 - \hat{Q}(t, \hat{E}|x, a) \right) \, dt + \hat{g}(x, a) \left( 1 - \hat{Q}(s, \hat{E}|x, a) \right) + \int_{[0,s]} \int_{\hat{E}} v(s - t, y) \hat{Q}(dt, dy|x, a).
\]

Moreover, for any $v \in \mathbb{M}$, $\varphi \in \Phi$ and $(s, x) \in [0, T] \times \hat{E}$, define

\[
\mathbb{T}^\varphi v(s, x) := \sum_{a \in A(x)} \varphi(a|s, x) \mathbb{T}^a v(s, x) \quad \text{and} \quad \mathbb{T} v(s, x) := \min_{a \in A(x)} \mathbb{T}^a v(s, x).
\]

**Remark 3.9** Here, we give an intuitive explanation of the operator $\mathbb{T}^a$. Given any planning horizon $s$ and initial state $x$, suppose that $U^f(s, x)$ denotes the expected cost under a deterministic stationary $f$. If an action $a := f(s, x) \in A(x)$ is chosen at $x$, then there are the following two cases.

The first case The first jump epoch $t$ occurs after the planning horizon $s$ (i.e., $t > s$), which means that the system stays at $x$ in $[0, s]$ with the probability $1 - \hat{Q}(s, \hat{E}|x, a)$. For this case, the incurred total cost in $[0, s]$ includes the running cost $c(x, a)s$ and the terminal cost $g(x, a)$, and then the expected cost is $(1 - \hat{Q}(s, \hat{E}|x, a)) \times (c(x, a)s + g(x, a))$.

The second case The first jump epoch $t$ occurs before the planning horizon $s$ (i.e., $t \leq s$) and the system jumps from $x$ to a state $y$ with probability $\hat{Q}(dt, dy|x, a)$. For this case, the remaining planning horizon becomes $s - t$ and the total cost in $[0, s]$ consists of the running cost $c(x, a)t$ and the subsequent cost for the rest planning horizon $U^f(s - t, y)$. Hence, the expected cost is $\int_{[0,s]} \int_{\hat{E}} (c(x, a)t + U^f(s - t, y)) \hat{Q}(dt, dy|x, a)$.

Considering the above two cases, we get

\[
U^f(s, x) = (1 - \hat{Q}(s, \hat{E}|x, a)) \times (c(x, a)s + g(x, a)) \\
+ \int_{[0,s]} \int_{\hat{E}} (c(x, a)t + U^f(s - t, y)) \hat{Q}(dt, dy|x, a) \\
= (1 - \hat{Q}(s, \hat{E}|x, a))g(x, a) + \int_{[0,s]} \int_{\hat{E}} U^f(s - t, y) \hat{Q}(dt, dy|x, a) \\
+c(x, a) \left[ s(1 - \hat{Q}(s, \hat{E}|x, a)) + \int_{[0,s]} t \hat{Q}(dt, \hat{E}|x, a) \right] \\
= \mathbb{T}^a U^f(s, x). \quad (3.10)
\]

In spirit of (3.10), we replace $U^f$ with any function $v$ and then obtain the definition of the operator $\mathbb{T}^a$.  

\[\odot\] Springer
To establish the iteration algorithm of $U^\pi$ and $U^\phi$, we define a function sequence \( \{U^\pi_n, n \geq -1\} \) as the following:

\[
U^\pi_{-1}(s, x) := 0, \quad U^\pi_n(s, x) := \sum_{m=0}^{n} \Phi^\pi_{(s, x)} \left[ \left((s - \hat{s}_m)^+ \land \hat{T}_{m+1}\right) \hat{c}(\hat{X}_m, A_m) + \mathbb{I}_{[\hat{s}_m, \hat{s}_{m+1})}(s) \hat{g}(\hat{X}_m, A_m) \right]
\]

for every \((s, x) \in [0, T] \times \hat{E}\) and \(n \geq 0\). It is a routine exercise to verify that \(U^\pi_n(s, x) \leq U^\pi_{n+1}(s, x)\) for each \(n \geq -1\). Moreover, (3.5) implies that \(\lim_{n \to \infty} U^\pi_n(s, x) = U^\pi(s, x)\). The following lemma is the basis of the optimality equation.

**Lemma 3.10** Suppose that Assumption 3.5 holds. Let \(\pi = \{\varphi_n, n \geq 0\} \in \Pi_{RM}\) be arbitrary.

(a) For each \(n \geq -1\), \(U^\pi_{n+1} = T^{\varphi_0} U^{(1)}_n\) and \(U^\pi = T^{\varphi_0} U^{(1)}\), where \((1)\pi := \{\varphi_n, n \geq 1\} \).

(b) In particular, for each \(\varphi \in \Phi\), \(U^\varphi_{n+1} = T^\varphi U^\varphi_n\) and \(U^\varphi = T^\varphi U^\varphi\).

**Proof** (a) First, since \(\pi\) is Markovian, the equalities (3.2)-(3.4) imply that

\[
\sum_{m=1}^{n+1} \Phi^\pi_{(s, x)} \left[ \left((s - \hat{s}_m)^+ \land \hat{T}_{m+1}\right) \hat{c}(\hat{X}_m, A_m) + \mathbb{I}_{[\hat{s}_m, \hat{s}_{m+1})}(s) \hat{g}(\hat{X}_m, A_m) \right]
\]

\[
= \sum_{m=1}^{n+1} \varphi_0(a|s, x) \int_{\hat{E}} \hat{Q}(dt, dy|x, a) \left[ \Phi^\pi_{(s, x)} \left((s - \hat{s}_m)^+ \land \hat{T}_{m+1}\right) \hat{c}(\hat{X}_m, A_m) + \mathbb{I}_{[\hat{s}_m, \hat{s}_{m+1})}(s) \hat{g}(\hat{X}_m, A_m) \right] \int_{\hat{E}} \hat{Q}(dt, dy|x, a)
\]

\[
= \sum_{a \in A(x)} \varphi_0(a|s, x) \int_{\hat{E}} \hat{Q}(dt, dy|x, a) \sum_{m=1}^{n+1} \Phi^\pi_{(s, x)} \left((s - \hat{s}_{m-1})^+ \land \hat{T}_m\right) \hat{c}(\hat{X}_{m-1}, A_{m-1}) + \mathbb{I}_{[\hat{s}_{m-1}, \hat{s}_m)}(s-t) \hat{g}(\hat{X}_{m-1}, A_{m-1})
\]

where the last equality is based on the fact that \(\mathbb{I}_{[\hat{s}_{m-1}, \hat{s}_m)}(s-t) = 0\) and \((s-t) - \hat{s}_{m-1} = 0\) when \(t > s\). Then, we have

\[
U^\pi_{n+1}(s, x) = \sum_{a \in A(x)} \varphi_0(a|s, x) \int_{\hat{E}} \hat{Q}(dt, dy|x, a) \hat{c}(\hat{X}_{m-1}, A_{m-1}) + \mathbb{I}_{[\hat{s}_{m-1}, \hat{s}_m)}(s-t) \hat{g}(\hat{X}_{m-1}, A_{m-1})
\]
Since $A(x)$ is finite, under Assumption 3.5, the monotone convergence theorem implies $U^\pi = \mathbb{T}^{\psi_0} U^{(1)^\pi}$.

(b) It can be obtained directly from part (a). \hfill \square

At the end of this section, we provide an iterative algorithm for computing the value function $U^*$ and give the optimality equation and the existence of $T$-optimal policies of the sMDPs with terminal costs.

**Theorem 3.11** Suppose that Assumption 3.5 holds. Then, the following statements hold.

(a) (Value iteration.) For every $n \geq 0$, let $U^*_n := 0$ and $U^*_n := \mathbb{T}U^*_n$. Then, we have $U^* = \lim_{n \to \infty} U^*_n \in \mathbb{M}$.

(b) (Optimality equation.) The value function $U^*$ is the minimum solution in $\mathbb{M}$ to the optimality equation $U^* = \mathbb{T}U^*$, i.e., if $u \in \mathbb{M}$ satisfies that $u = \mathbb{T}u$, then $u \geq U^*$.

(c) ($T$-optimal policy.) There exists a function $f^* \in \mathbb{F}$ such that $U^* = \mathbb{T}f^* U^*$. Hence, the deterministic stationary policy $f^*$ is $T$-optimal.

**Proof** (a) Since $\hat{c}(x, a)$ and $\hat{g}(x, a)$ are nonnegative, $U^* = 0$ and $\mathbb{T}$ is nondecreasing, we obtain $U^*_{n+1}(s, x) \geq U^*_n(s, x)$ and $U^*_n \in \mathbb{M}$ for all $(s, x) \in [0, T] \times \tilde{E}$ and $n \geq 0$. Therefore, $u^* := \lim_{n \to \infty} U^*_n \in \mathbb{M}$. To prove part (a), it remains to establish that $u^* = U^*$. Next, we prove $u^* \leq U^*$ and $U^* \geq u^*$.

In order to show $u^* \leq U^*$, we just need to verify that for each $\pi \in \Pi_{RM}$,

$$U^*_{n+1} \leq U^*_n, \quad \forall n \geq -1. \quad (3.11)$$

It is obviously true for $n = -1$. Suppose that $U^*_{n+1} \leq U^*_{n} \pi$ for some $n \geq -1$ and all $\pi \in \Pi_{RM}$. Then, for any $\pi = \{\psi_n, n \geq 0\} \in \Pi_{RM}$, by part (a) of Lemma 3.10, it holds that

$$U^*_{n+1} = \mathbb{T}^{\psi_0} U^{(1)^* \pi} \geq \mathbb{T}^{\psi_0} U^*_{n+1} \geq \mathbb{T}U^*_n = U^*_n + 2,$$

where $(1)^\pi = \{\psi_n, n \geq 1\} \in \Pi_{RM}$ and the second and third inequalities follow from the inductive hypothesis and the definitions of $\mathbb{T}$ and $\mathbb{T}^{\psi_0}$. By induction, the inequality (3.11) holds.
We now show that $u^* \geq U^*$. Fix any $(s, x) \in [0, T] \times \hat{E}$. Since $A(x)$ is finite, there exists an $a^*_n(s, x)$ satisfying that $\mathbb{T}^{a^*_n(s, x)}U^*_n(s, x) = \mathbb{T}U^*_n(s, x) = U^*_{n+1}(s, x)$. Again, the finiteness of $A(x)$ implies that there exists a subsequence $\{n_k\}$ and $a^*(s, x) \in A(x)$ such that $a^*_n(s, x) = a^*(s, x)$ for all $n_k$. Hence, $U^*_{n_k+1}(s, x) = \mathbb{T}^{a^*(s, x)}U^*_n(s, x)$. Passing the limit $n_k \to \infty$, the monotone convergence theorem ensures that $u^*(s, x) = \mathbb{T}^{a^*(s, x)}u^*(s, x)$, which implies that $u^*(s, x) \geq \mathbb{T}u^*(s, x)$. By the arbitrariness of $(s, x)$, we have $u^* \geq \mathbb{T}u^*$. On the other hand, the finiteness of $A(x)$ and the measurable selection theorem (see [16, Proposition D.5]) ensure that there exists $f^* \in \mathcal{F}$ such that $\mathbb{T}f^* u^* = \mathbb{T}u^* \leq u^*$.

Moreover, since $u^* \geq 0$, by induction it holds that $u^* \geq \mathbb{T}^{f^*} U^*_n = U^*_{n+1}$ for all $n \geq 1$. Hence, we have
\[
 u^* \geq \lim_{n \to \infty} U^*_n = U^{f^*} \geq U^*. \tag{3.12}
\]

To sum up, we have proved $u^* = U^*$.

(b) For $\pi = \{\varphi_n, n \geq 0\} \in \Pi_{RM}$, by part (a) of Lemma 3.10, it holds that
\[
 U^\pi = \mathbb{T}^{\varphi_0} U^{(1)\pi} \geq \mathbb{T}^{\varphi_0} U^* \geq \mathbb{T}U^*.
\]

Then, the arbitrariness of $\pi$ implies that $U^* \geq \mathbb{T}U^*$. On the other hand,
\[
 U^*_{n+1}(s, x) = \mathbb{T}U^*_n(s, x) \leq \mathbb{T}U^*_n(s, x), \quad \forall (s, x) \in [0, T] \times \hat{E}, \quad a \in A(x), \quad n \geq 0.
\]

Hence, by the monotone convergence theorem, we obtain that $U^*(s, x) \leq \mathbb{T}U^*(s, x)$, and then $U^*(s, x) \leq \mathbb{T}U^*(s, x)$. Therefore, we obtain $U^* = \mathbb{T}U^*$.

Let $u \in \mathbb{M}$ be an arbitrary solution to the equation $u = \mathbb{T}u$. Denote $\mathbb{T}^{n+1}v := \mathbb{T}(\mathbb{T}^nv)$ for any $v \in \mathbb{M}$ and $n \geq 1$. Since $u \geq 0 = U^*_0$ and $u = \mathbb{T}^nu$, it follows from part (a) that
\[
 u = \lim_{n \to \infty} \mathbb{T}^{n+1}u \geq \lim_{n \to \infty} \mathbb{T}^{n+1}U^*_0 = \lim_{n \to \infty} U^*_n = U^*.
\]

This means that $U^*$ is the minimum solution in $\mathbb{M}$ to the optimality equation.

(c) According to the proof of part (a), there exists $f^* \in \mathcal{F}$ such that $\mathbb{T}U^* = \mathbb{T}f^* U^*$. Therefore, part (b) gives $U^* = \mathbb{T}f^* U^*$. Hence, $f^*$ is a $T$-optimal policy by part (a) and (3.12). $\square$

**Remark 3.12** In particular, the above results degenerate the previous results studied in [17, Theorem 3.1 and Theorem 3.2] when $\hat{E}$ is denumerable and $\hat{g} \equiv 0$.

### 4 The Existence and Computation of the $T$-optimal Stopping Times

In this section, we consider the optimal stopping problems on sMPs in Sect. 2. The main method is transforming the optimal stopping problems on sMPs into the equivalent sMDPs and then using the conclusions (Theorem 3.11) of sMDPs introduced in Sect. 3.
First, for each stopping time $\tau$ and $s \in [0, T]$, we will show that there is a policy $\pi_\tau$ such that the $s$-horizon expected cost of the policy $\pi_\tau$ in sMDPs is equal to that of $\tau$ in sMPs, i.e., $V^\tau(s, x) = U_{\pi_\tau}(s, x)$. Hence, we can analyze the value function $V^\tau$ and the $T$-optimal stopping time $\tau^*$ of optimal stopping problems on sMPs through the corresponding sMDPs.

Intuitively, in the optimal stopping problems, “continue” or “stop” can be considered as a special action. This intuition gives us an idea to construct the sMDPs. The constructions of the sMDPs are given as follows:

$${\hat{E}, A, (A(x), x \in {\hat{E}}), {\hat{Q}}_T(\cdot, \cdot|x, a), {\hat{c}}(x, a), {\hat{g}}(x, a)}.$$  \tag{4.1}

The state space $\hat{E} := E \cup \{\Delta\}$ includes the state space $E$ of sMPs and a virtual state $\{\Delta\}$. The set of admissible actions $A(x)$ at state $x \in \hat{E}$ is defined as

$$A(x) := \begin{cases} \{0, 1\}, & x \in E; \\ \{1\}, & x = \Delta, \end{cases}$$

where the action 0 means continuation and 1 means stop. The action space $A = \{0, 1\}$ is finite, and the set of admissible state-action pairs $K = (E \times A) \cup \{(\Delta, 1)\}$ is a Borel subset of $\hat{E} \times A$. For each $t \geq 0$ and $B \in \mathcal{B}(\hat{E})$, the semi-Markov kernel $\hat{Q}_T(t, B|x, a)$ is given by

$$\hat{Q}_T(t, B|x, a) := \begin{cases} Q(t, B \setminus \{\Delta\}|x), & x \in E, a = 0; \\ 1_{[T+1, +\infty)}(t)\delta_\Delta(B), & x \in \hat{E}, a = 1, \end{cases} \tag{4.2}$$

where $Q(t, B|x)$ is the kernel of sMPs. Finally, the cost rate function and terminal cost function are

$$\hat{c}(x, a) := \begin{cases} c(x), & x \in E, a = 0; \\ 0, & \text{otherwise}, \end{cases} \tag{4.3}$$

$$\hat{g}(x, a) := \begin{cases} g(x), & x \in E, a = 1; \\ 0, & \text{otherwise}, \end{cases} \tag{4.4}$$

where $c$ and $g$ are given in (2.3). Since $c$ and $g$ are measurable on $E$, $\hat{c}(x, a)$ and $\hat{g}(x, a)$ are also measurable on $K$. Firstly, we give a lemma to show that the sMDPs given above satisfy Assumption 3.5.

**Lemma 4.1** Suppose that Assumption 2.2 holds. For the sMDPs introduced in (4.1), Assumption 3.5 holds.

**Proof** Under Assumption 2.2, there are $\delta > 0$ and $\alpha > 0$ satisfying that (2.1). Let $\hat{\delta} = \min\{\delta, 1/2\}$, and then

$$\hat{Q}_T(\hat{\delta}, \hat{E}|x, a) = \begin{cases} Q(\hat{\delta}, E|x) \leq Q(\delta, E|x) \leq 1 - \alpha, & x \in E, a = 0; \\ 1_{[T+1, +\infty)}(\hat{\delta}) = 0 \leq 1 - \alpha, & x \in \hat{E}, a = 1. \end{cases}$$

According to the result in [17, Proposition 2.1], Assumption 3.5 holds. \qed
In the next step, we will give a relationship between the stopping times \( \tau \in \Gamma \) in sMPs and the policies \( \pi \in \Pi_{DH} \) in sMDPs defined in (4.1). For each \( n \geq 0 \) and the history \( h_n \in H_n \) of sMPs up to the \( n \)th jump epoch, let

\[
\hat{h}_n^0 = (x_0, 0, t_1, x_1, \ldots, 0, t_n, x_n) \in \hat{H}_n.
\]

(4.5)

The action 0 (means continuation) added to equation (4.5) indicates that the system has been running incessantly before the \( n \)th jump epoch. By the definition of \( \hat{h}_n^0 \), it holds that

\[
\hat{C}^0 := \{ \hat{h}_n^0 | h_n \in C \} \in B(\hat{H}_n) \quad \forall C \in B(H_n).
\]

(4.6)

In particular, \( \hat{H}_n^0 = E \times (\{0\} \times \mathbb{R}_+ \times E)^n \in B(\hat{H}_n) \).

**Definition 4.2** Given any deterministic policy \( \pi = \{d_n, n \geq 0\} \in \Pi_{DH} \) defined in Definition 3.4 (c) and \( s \in \mathbb{R} \), define

\[
\tau^s_\pi(\omega) := \inf \left\{ n \in \mathbb{N} \mid d_n(s - S_n(\omega), \hat{Y}_n^0(\omega)) = 1 \right\},
\]

\( \omega = (x_0, t_1, \ldots, x_n, t_{n+1}, \ldots) \in \Omega \),

where \( \inf \emptyset := +\infty \) and \( \hat{Y}_n^0 := (X_0, 0, T_1, X_1, \ldots, 0, T_n, X_n) \). Then, \( \tau^s_\pi \) is called the stopping time induced by the policy \( \pi \) and \( s \).

**Lemma 4.3** For each deterministic policy \( \pi = \{d_n, n \geq 0\} \) and \( s \in \mathbb{R} \), the induced stopping time \( \tau^s_\pi \) is a stopping time.

**Proof** Note that for each \( n \geq 0 \), the random variables \( \hat{Y}_n^0, S_n \) and the function \( d_n \) are measurable in their corresponding spaces. Hence, we have

\[
\{ \tau^s_\pi = n \} = \left( \bigcap_{k=0}^{n-1} \{ d_k((s - S_k, \hat{Y}_k^0) = 0) \} \right) \cap \{ d_n((s - S_n, \hat{Y}_n^0) = 1 \} \in \sigma(Y_n) = \mathcal{F}_n,
\]

which implies that \( \tau^s_\pi \) is a stopping time. \( \square \)

We introduce a subclass \( \Pi_{DH}^0 \) of \( \Pi_{DH} \) by

\[
\Pi_{DH}^0 := \{ \pi = \{d_n, n \geq 0\} \in \Pi_{DH} | d_n(0, \hat{h}_n) = 0, \forall n \geq 0, \hat{h}_n \in \hat{H}_n^0 \}.
\]

We now give a key theorem to establish the relationship between the \( s \)-horizon expected cost of polices in \( \Pi_{DH}^0 \) and that of the stopping times induced by these polices and \( s \).

**Theorem 4.4** Suppose that Assumption 2.2 holds. For any \( \pi = \{d_n, n \geq 0\} \in \Pi_{DH}^0 \), it holds that

\[
U^\pi(s, x) = V^{\tau^s_\pi}(s, x) \quad \forall x \in E, s \in [0, T],
\]

\( \square \) Springer
where $\tau^s_\pi$ is the stopping time induced by $\pi$ and $s$ and $V^{\tau^s_\pi}(s, x)$ is the $s$-horizon expected cost of $\tau^s_\pi$.

**Proof** For each $\hat{\omega} = (x_0, a_0, t_1, \ldots, x_n, a_n, t_{n+1}, \ldots) \in \hat{\Omega}$, recall that $\hat{S}_n(\hat{\omega}) = \sum_{k=1}^{n} t_k$ and

$$\hat{Y}_n(\hat{\omega}) = (x_0, a_0, t_1, x_1, \ldots, a_{n-1}, t_n, x_n).$$

We define $C_n (n \geq 0)$ and $C$, the subsets of $\hat{\Omega}$, as

$$C_n := \left\{ \hat{\omega} \in \hat{\Omega} : \inf \left\{ k \in \mathbb{N} : d_k (s - \hat{S}_k(\hat{\omega}), \hat{Y}_k(\hat{\omega})) = 1 \right\} = n \right\}, \quad n \geq 0;$$

$$C := \left\{ \hat{\omega} \in \hat{\Omega} : d_k (s - \hat{S}_k(\hat{\omega}), \hat{Y}_k(\hat{\omega})) = 0, \forall k \geq 0 \right\}.$$

It can be verified that $\{C, C_n, n \geq 0\}$ is a partition of $\hat{\Omega}$ and

$$1_{C_n} = \prod_{k=0}^{n-1} 1_{[0]} (d_k (s - \hat{S}_k, \hat{Y}_k)) \times 1_{[1]} (d_n (s - \hat{S}_n, \hat{Y}_n)).$$

Hence, the monotone convergence theorem implies that

$$U^\pi (s, x) = \sum_{n=0}^{\infty} E^\pi (s, x) \left[ 1_{C_n} \left( \int_0^{\tau^s_\pi} \hat{c}(\hat{X}(t), A(t))dt + \hat{g}(\hat{X}(s), A(s)) \right) \right]$$

$$+ E^\pi (s, x) \left[ 1_C \left( \int_0^{\tau^s_\pi} \hat{c}(\hat{X}(t), A(t))dt + \hat{g}(\hat{X}(s), A(s)) \right) \right], \quad (4.7)$$

Noting that $\pi \in \Pi^0_{DH}$, the definition of $C_n$ and (3.3) give

$$\hat{E}^\pi_{(s, x)} [A_k = 1 | C_n] = \hat{E}^\pi_{(s, x)} \left[ d_k (s - \hat{S}_k, \hat{Y}_k) = 1 | C_n \right] = \begin{cases} 0, & k < n; \\ 1, & k = n, \end{cases} \quad (4.8)$$

which, together with $\lim_{t \to \infty} \hat{Q}_T (t, \Delta | x, 1) = 1$ and $A(\Delta) = \{1\}$, implies that $\hat{E}^\pi_{(s, x)} (\hat{X}_m = \Delta | C_n) = 1$ for all $m > n$. Thus, by Lemma 4.1, (3.5), (4.3), and (4.4) give

$$\hat{E}^\pi_{(s, x)} \left[ 1_{C_n} \left( \int_0^{\tau^s_\pi} \hat{c}(\hat{X}(t), A(t))dt + \hat{g}(\hat{X}(s), A(s)) \right) \right]$$

$$= \sum_{m=0}^{n-1} \hat{E}^\pi_{(s, x)} \left[ \hat{c}(\hat{X}_m, A_m) ((s - \hat{S}_m)^+ \wedge \hat{T}_m+1) 1_{C_n} \right] + \hat{E}^\pi_{(s, x)} \left[ 1_{C_n} 1_{[\hat{S}_n, \hat{T}_{n+1}]} (s) \hat{g}(\hat{X}_n, 1) \right], \quad (4.9)$$
And then, for each \( m < n \), using

\[
\mathbb{1}_{\tau^+ = n} = \prod_{k=0}^{n-1} \mathbb{1}_{[0]}(d_k(s - S_k, \hat{Y}^0_k)) \times \mathbb{1}_{[1]}(d_n(s - S_n, \hat{Y}^0_n)) \tag{4.10}
\]

and \( \mathcal{Q}_T(\cdot, B|0, \omega) = Q(\cdot, B \setminus \{\Delta\}|x) \) for all \( x \in E \) and \( B \in \mathcal{B}(\hat{E}) \), it holds that

\[
\hat{\mathcal{P}}_{(s,x)} \left[ \hat{c}(\hat{X}_m, A_m)((s - \hat{S}_m)^+ \land \hat{T}_{m+1}) \mathbb{1}_C \right] = \int_{\hat{E}} \delta_x(dx_0) \int_{\hat{E}} \int_{\mathbb{R}^+} \hat{Q}_T(dr_1, dx_1|x_0, d_0(s, \hat{h}_0)) \int_{\hat{E}} \int_{\mathbb{R}^+} \hat{Q}_T(dr_2, dx_2|x_1, d_0(s - s_1, \hat{h}_1)) \cdots \int_{\hat{E}} \int_{\mathbb{R}^+} \hat{Q}_T(dr_n, dx_n|x_{n-1}, d_{n-1}(s - s_{n-1}, \hat{h}_{n-1})) \hat{c}(x_m, d_m(s - s_m, \hat{h}_m))
\]

\[
= \int_{\hat{E}} \delta_x(dx_0) \int_{\hat{E}} \int_{\mathbb{R}^+} Q(dr_1, dx_1|x_0) \cdots \int_{\hat{E}} \int_{\mathbb{R}^+} Q(dr_n, dx_n|x_{n-1}) c(x_m)
\]

\[
((s - s_m)^+ \land \tau_m+1) \prod_{k=0}^{n-1} \mathbb{1}_{[0]}(d_k(s - s_k, \hat{h}_k)) \cdot \mathbb{1}_{[1]}(d_n(s - s_n, \hat{h}_n))
\]

\[
= \mathbb{E}_x \left[ ((s - S_m)^+ \land T_{m+1})c(X_m) \mathbb{1}_{\tau^+ = n} \right], \tag{4.11}
\]

where \( s_k = \sum_{i=t_i}^{t_{i+1}} t_i, \hat{h}_0 = x_0, \hat{h}_{k+1} = (\hat{h}_k, d_k(s - s_k, \hat{h}_k), t_{k+1}, x_{k+1}), h_k = (x_0, t_1, x_1, \ldots, t_k, x_k) \) and \( \hat{h}_k \) is defined in (4.5). For each \( \hat{\omega} \in C_n, \pi \in \Pi_{HD}^0 \) gives \( \hat{Y}_n(\hat{\omega}) \in \hat{H}_n^0 \) and \( d_n(0, \hat{Y}_n(\hat{\omega})) = 0 \). Furthermore, by \( d_n(s - \hat{S}_n(\hat{\omega}), \hat{Y}_n(\hat{\omega})) = 1 \) for all \( \hat{\omega} \in C_n \), we have \( C_n \cap \{s = \hat{S}_n\} = \emptyset \). Therefore, we have \( \mathbb{1}_{C_n} \mathbb{1}_{\tau^+ = n} = \mathbb{1}_{C_n} \mathbb{1}_{\hat{S}_n, \hat{S}_{n+1}}(s) \). Thus, (4.2) and (4.4) show

\[
\hat{\mathcal{P}}_{(s,x)} \left[ \mathbb{1}_{C_n} \mathbb{1}_{\hat{S}_n, \hat{S}_{n+1}}(s) \hat{g}(\hat{X}_n, 1) \right] = \int_{\hat{E}} \delta_x(dx_0) \int_{\hat{E}} \int_{\mathbb{R}^+} \hat{Q}_T(dr_1, dx_1|x_0, d_0(s, \hat{h}_0)) \int_{\hat{E}} \int_{\mathbb{R}^+} \hat{Q}_T(dr_2, dx_2|x_1, d_0(s - s_1, \hat{h}_1)) \cdots \int_{\hat{E}} \int_{\mathbb{R}^+} \hat{Q}_T(dr_n, dx_n|x_{n+1}, d_n(s - s_n, \hat{h}_n)) \hat{g}(x_n, 1) \mathbb{1}_{(s_n, s_{n+1}+t_n+1)}(s)
\]

\( \copyright \) Springer
\[
\prod_{k=0}^{n-1} \mathbb{1}_{\{0\}}(d_k(s - s_k, \hat{h}_k)) \mathbb{1}_{\{1\}}(d_n(s - s_n, \hat{h}_n))
\]

\[
\int_{E} \delta_{X}(dx_0) \int_{E} \int_{\mathbb{R}_+^2} Q(dt_1, dx_1|x_0)
\]

\[
\cdots \int_{E} \int_{\mathbb{R}_+^2} Q(dt_n, dx_n|x_{n-1}) g(x_n) \mathbb{1}_{(s_n, s_n+T+1)}(s)
\]

\[
\times \prod_{k=0}^{n-1} \mathbb{1}_{\{0\}}(d_k(s - s_k, \hat{h}_k^0)) \cdot \mathbb{1}_{\{1\}}(d_n(s - s_n, \hat{h}_n^0))
\]

(by \(\hat{Q}_T(t, \hat{E}|x_n, 1) = \mathbb{1}_{[T+1, \infty)}(t)\))

\[
\mathbb{E}_x \left[ \mathbb{1}_{(S_n, +\infty)}(s) g(X_n) \mathbb{1}_{\{\tau_n^s = n\}} \right]. \quad (4.12)
\]

Moreover, by the definition of \(R_{\tau_n^s}^s\) given in (2.3), (4.9), (4.11), and (4.12), we have

\[
\mathbb{E}_x \left[ \mathbb{1}_{\{\tau_n^s = n\}} R_{\tau_n^s}^s \right] = \mathbb{E}_x \left[ \mathbb{1}_{\{S_n < s\}} \mathbb{1}_{\{\tau_n^s = n\}} \left( \int_0^{S_n} c(X(t))dt + g(X(S_n)) \right) \right]
\]

\[
+ \mathbb{E}_x \left[ \mathbb{1}_{\{S_n \geq s\}} \mathbb{1}_{\{\tau_n^s = n\}} \int_0^s c(X(t))dt \right]
\]

\[
= \mathbb{E}_x \left[ \mathbb{1}_{\{S_n < s\}} \mathbb{1}_{\{\tau_n^s = n\}} g(X_n) \right]
\]

\[
+ \mathbb{E}_x \left[ \mathbb{1}_{\{S_n < s\}} \mathbb{1}_{\{\tau_n^s = n\}} \sum_{m=0}^{n-1} c(X_m)((s - S_m)^+ \land T_{m+1}) \right]
\]

\[
+ \mathbb{E}_x \left[ \mathbb{1}_{\{S_n \geq s\}} \mathbb{1}_{\{\tau_n^s = n\}} \sum_{m=0}^{n-1} c(X_m)((s - S_m)^+ \land T_{m+1}) \right]
\]

\[
= \hat{\mathbb{E}}_{(s, x)}^{\tau} \left[ \mathbb{1}_{C_n} \left( \int_0^s \hat{c}(\hat{X}(t), A(t))dt + \hat{g}(\hat{X}(s), A(s)) \right) \right]. \quad (4.13)
\]

Next, we calculate the second item of (4.7). Since \(\{C, C_n, n \geq 0\}\) is a partition of \(\hat{\Omega}\), we obtain that

\[
\mathbb{1}_C = \mathbb{1}_C \times \prod_{m=0}^{k} (1 - \mathbb{1}_{C_m}) = (1 - \sum_{n=0}^{+\infty} \mathbb{1}_{C_n}) \times \prod_{m=0}^{k} (1 - \mathbb{1}_{C_m})
\]

\[
= \prod_{m=0}^{k} (1 - \mathbb{1}_{C_m}) - \sum_{n=k+1}^{+\infty} \mathbb{1}_{C_n} \forall k \geq 0,
\]

which, combining with (3.5), implies that

\[
\hat{\mathbb{E}}_{(s, x)}^{\tau} \left[ \mathbb{1}_C \left( \int_0^s \hat{c}(\hat{X}(t), A(t))dt + \hat{g}(\hat{X}(s), A(s)) \right) \right]
\]
\[
= \sum_{k=0}^{\infty} \left\{ \hat{E}_{\pi}^{\tau_{s,\pi}(s)} \left[ \prod_{m=0}^{k} (1 - \mathbb{1}_{C_m})( (s - \hat{S}_k)^+ \wedge \hat{T}_{k+1}) \hat{c}(\hat{X}_k, A_k) \right] \\
+ \mathbb{1}_{[\hat{S}_k, \hat{S}_{k+1})}(s) \hat{g}(\hat{X}_k, A_k) \right\] \\
- \sum_{n=k+1}^{\infty} \mathbb{1}_{[\tau_{s,\pi}(s) = +\infty]} \hat{E}_{\pi}^{\tau_{s,\pi}(s)} \left[ \prod_{m=0}^{n} (1 - \mathbb{1}_{C_m})( (s - \hat{S}_m)^+ \wedge \hat{T}_{m+1}) \hat{c}(\hat{X}_k, A_k) \\
+ \mathbb{1}_{[\hat{S}_k, \hat{S}_{k+1})}(s) \hat{g}(\hat{X}_k, A_k) \right] \right\}.
\]

Firstly, using that \(\hat{g}(x, 0) = 0\) for all \(x \in E\), (4.8) and (4.11), we obtain that for each \(n > k\),
\[
\hat{E}_{\pi}^{\tau_{s,\pi}(s)} \left[ \prod_{m=0}^{k} (1 - \mathbb{1}_{C_m})( (s - \hat{S}_k)^+ \wedge \hat{T}_{k+1}) \hat{c}(\hat{X}_k, A_k) \\
+ \mathbb{1}_{[\hat{S}_k, \hat{S}_{k+1})}(s) \hat{g}(\hat{X}_k, A_k) \right] \\
= \mathbb{E}_x \left[ (s - S_m)^+ \wedge T_{m+1} \hat{c}(X_m) \mathbb{1}_{[\tau_{s,\pi}(s) = n]} \right].
\]

And for any \(k \geq 0\), by the same methods of (4.11)
\[
\hat{E}_{\pi}^{\tau_{s,\pi}(s)} \left[ \prod_{m=0}^{k} (1 - \mathbb{1}_{C_m})( (s - \hat{S}_m)^+ \wedge \hat{T}_{m+1}) \hat{c}(\hat{X}_k, A_k) \\
+ \mathbb{1}_{[\hat{S}_k, \hat{S}_{k+1})}(s) \hat{g}(\hat{X}_k, A_k) \right] \\
= \mathbb{E}_x \left[ \prod_{m=0}^{k} (1 - \mathbb{1}_{[\tau_{s,\pi}(s) = m]})( (s - S_m)^+ \wedge T_{m+1}) \hat{c}(X_k) \right].
\]

Therefore, we obtain that
\[
\mathbb{E}_x \left[ \mathbb{1}_{[\tau_{s,\pi}(s) = +\infty]} R_{\tau_{s,\pi}}(s) \right] = \hat{E}_{\pi}^{\tau_{s,\pi}(s)} \left[ \mathbb{1}_C \left( \int_0^s \hat{c}(\hat{X}(t), A(t)) dt + \hat{g}(\hat{X}(s), A(s)) \right) \right],
\]

which, together with (4.7) and (4.13), shows that
\[
U_{\pi}(s, x) = \sum_{n=0}^{\infty} \mathbb{E}_x \left[ \mathbb{1}_{[\tau_{s,\pi}(s) = n]} R_{\tau_{s,\pi}}(s) \right] + \mathbb{E}_x \left[ \mathbb{1}_{[\tau_{s,\pi}(s) = +\infty]} R_{\tau_{s,\pi}}(s) \right] \\
= V_{\tau_{s,\pi}(s)}(s, x) \ \forall x \in E, s \in [0, T].
\]

The proof of Theorem 4.4 is completed. \(\square\)

For any \(s \in [0, T]\), Definition 4.2, Lemma 4.3, and Theorem 4.4 say that for each policy \(\pi \in \Pi_{DH}^0\), we can construct a stopping time \(\tau_{s,\pi}(s)\) such that their \(s\)-horizon expected costs are equal. On the other hand, for each stopping time we also can
construct a policy which satisfies this condition, see Definition 4.5, Lemma 4.6, and Theorem 4.7.

**Definition 4.5** Given any stopping time $\tau \in \Gamma$ and $n \geq 0$, let

$$B^\tau_n := \{ Y_n(\omega) : \omega = (x_0, t_1, x_1, \ldots, t_k, x_k, \ldots) \in \{ \tau = n \} \},$$

where $Y_n(\omega) = (x_0, t_1, x_1, \ldots, t_n, x_n)$. For each $\hat{h}_n = (x_0, a_0, t_1, x_1, \ldots, a_{n-1}, t_n, x_n) \in \hat{H}_n$, $s \in \mathbb{R}$, define

$$d^\tau_n(s, \hat{h}_n) := \begin{cases} 1_{B^\tau_n}(x_0, t_1, x_1, \ldots, t_n, x_n)1_{(0, \infty)}(s), & \hat{h}_n \in \hat{H}_n^0, \\ 1, & \hat{h}_n \in \hat{H}_n \setminus \hat{H}_n^0, \end{cases}$$

where $\hat{H}_n^0 = E \times ((0) \times \mathbb{R}_+ \times E)^n$. $\pi_\tau := \{ d^\tau_n, n \geq 0 \}$ is called the policy induced by $\tau$.

**Lemma 4.6** For each stopping time $\tau \in \Gamma$, the policy $\pi_\tau$ induced by $\tau$ is a deterministic policy of the corresponding sMDPs, more specifically $\pi_\tau \in \Pi_{DH}^0 \subset \Pi_{DH}$.

**Proof** By Definition 4.5, it can be verified that $d^\tau_n(s, \hat{h}_n) \in A(x_n)$. Then, we just need to consider the measurability. Noting that $\{ \tau = n \} \in \mathcal{F}_n = \sigma(Y_n)$, we obtain $B^\tau_n \in \mathcal{B}(H_n)$. Thus, we have

$$\{ (s, \hat{h}_n) \in \mathbb{R} \times \hat{H}_n | d^\tau_n(s, \hat{h}_n) = 0 \} = \left( (0, +\infty) \times \hat{C}^0_n, \tau \right)$$

$$\cup (-\infty, 0] \times \hat{H}_n^0) \in \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\hat{H}_n),$$

where $C_{n, \tau} = H_n \setminus B^\tau_n \in \mathcal{B}(H_n)$ and $\hat{C}^0_n, \tau$ is defined in (4.6). Hence, $\pi_\tau := \{ d^\tau_n, n \geq 0 \}$ is a deterministic policy of the corresponding sMDPs. Furthermore, for each $\hat{h}_n \in \hat{H}_n^0$, it holds that $d^\tau_n(0, \hat{h}_n) = 0$, which implies that $\pi_\tau \in \Pi_{DH}^0$. 

**Theorem 4.7** Suppose that Assumption 2.2 holds. For each stopping time $\tau \in \Gamma$, let $\pi_\tau := \{ d^\tau_n, n \geq 0 \}$ be the policy induced by $\tau$. Then,

$$V^\tau(s, x) = U^{\pi_\tau}(s, x) = V^{\tau^s_{\pi_\tau}}(s, x) \quad \forall (s, x) \in [0, T] \times E,$$

where $\tau^s_{\pi_\tau}$ is the stopping time induced by $\pi_\tau$ and $s$.

**Proof** To prove (4.14), by Theorem 4.4, it suffices to show that $V^\tau(s, x) = V^{\tau^s_{\pi_\tau}}(s, x)$. By (4.10) and the definition of $d^\tau_n$, we have

$$1_{\{ \tau^s_{\pi_\tau} = n \}} = \prod_{k=0}^{n-1} 1_{[0]}(d^\tau_k(s - S_k, \hat{Y}_k^0)) \cdot 1_{[1]}(d^\tau_n(s - S_n, \hat{Y}_n^0)) = 1_{[0, s]}(S_n)1_{[\tau = n]},$$

(4.15)
\[
\hat{Y}_k^0 = (X_0, 0, T_1, X_1, \ldots, 0, T_k, X_k). \text{ Moreover, using the definition of } \tau_{\pi_T}^s \text{ again,}
\]

\[
1_{\{\tau_{\pi_T}^s = \infty\}} = \prod_{n=0}^{\infty} \left( 1 - 1_{\{\tau = n\}} 1_{(0, \infty)}(s - S_n) \right)
\]

\[
= 1_{\{\tau = \infty\}} \prod_{n=0}^{\infty} \left( 1 - 1_{\{\tau = n\}} 1_{(0, \infty)}(s - S_n) \right)
\]

\[
+ 1_{\{\tau \neq \infty\}} \prod_{n=0}^{\infty} \left( 1 - 1_{\{\tau = n\}} 1_{(0, \infty)}(s - S_n) \right)
\]

\[
= 1_{\{\tau = \infty\}} + \sum_{k=0}^{\infty} 1_{\{\tau = k\}} 1_{[s, \infty)}(S_k).
\] (4.16)

According to the definition of \( R_T^s \) given in (2.3), (4.15), and (4.16), we have that

\[
V^{\tau_{\pi_T}}(s, x) = \sum_{n=0}^{\infty} \mathbb{E}_x \left[ 1_{\{\tau_{\pi_T}^s = n\}} R_T^n \right] + \mathbb{E}_x \left[ 1_{\{\tau_{\pi_T}^s = \infty\}} \int_0^s c(X(t))dt \right]
\]

\[
= \sum_{n=0}^{\infty} \mathbb{E}_x \left[ 1_{[0, s)}(S_n) 1_{\{\tau = n\}} R_T^n \right]
\]

\[
+ \mathbb{E}_x \left[ 1_{\{\tau = \infty\}} + \sum_{n=0}^{\infty} 1_{\{\tau = n\}} 1_{[s, \infty)}(S_n) \right] \int_0^s c(X(t))dt \right]
\]

\[
= \sum_{n=0}^{\infty} \left( \mathbb{E}_x \left[ 1_{[0, s)}(S_n) 1_{\{\tau = n\}} R_T^n \right] \right)
\]

\[
+ \mathbb{E}_x \left[ 1_{\{\tau = n\}} 1_{[s, \infty)}(S_n) R_T^n \right] + \mathbb{E}_x \left[ 1_{\{\tau = \infty\}} R_T^s \right]
\]

\[
= \mathbb{V}_T(s, x),
\]

which is the desired result. \( \Box \)

We next establish the existence and computation of a finite \( T \)-optimal stopping time as well as the value function \( V^* \). For any \( x \in E \) and \( s \in [0, T] \), by the definition of \( T^a \) and \( A(x) = \{0, 1\} \), we have that

1. \( T^0 U^{\pi_T}(s, x) \)

\[
= \hat{c}(x, 0) \int_0^s (1 - \hat{Q}_T(t, \hat{E}|x, 0))dt + \hat{g}(x, 0)(1 - \hat{Q}_T(s, \hat{E}|x, 0)
\]

\[
+ \int_{[0, s]} \int_{\hat{E}} U^{\pi_T}(s - t, y) \hat{Q}_T(dt, dy|x, 0)
\]

\[
= c(x) \int_0^s (1 - Q(t, E|x))dt
\]
We define the policy \( x \) as deterministic stationary policy. Moreover, for any \( \{ s \} \), which implies

\[
\mathbb{E}(\tau^{\ast}(s, x)) = \frac{g(x)}{Q_T(s, x)} + \int_0^s \mathbb{E}(\tau^{\ast}(s, x)) dQ_T(s, x) \quad \text{for} \quad s \in [0, T].
\]

which mean that for each \( (s, x) \in [0, T] \times E \),

\[
\mathbb{T}U^{\pi^{\ast}}(s, x) = \min \left\{ \mathbb{T}^0 U^{\pi^{\ast}}(s, x), \mathbb{T}^1 U^{\pi^{\ast}}(s, x) \right\}
\]

\[
= \left\{ c(x) \int_0^T (1 - Q(t, E|x)) dt + \int_E \mathbb{T}^1 U^{\pi^{\ast}}(s, x) Q(dt, dy|x), g(x) \right\}. \quad (4.17)
\]

In spirit of Theorem 3.11, Theorem 4.7, and (4.17), we introduce an operator \( \mathbb{G} : \mathcal{M} \to \mathcal{M} \) as

\[
\mathbb{G} u(s, x) = \min \left\{ c(x) \int_0^T (1 - Q(t, E|x)) dt + \int_E u(s - t, x) Q(dt, dy|x), g(x) \right\}, u \in \mathcal{M}, (s, x) \in [0, T] \times E,
\]

where \( \mathcal{M} \) denotes the set of Borel measurable functions \( u : [0, T] \times E \to [0, \infty] \).

**Theorem 4.8 (Value iteration.)** Suppose that Assumption 2.2 holds. For any \( n \geq 0 \), let \( V_0^{\ast} \equiv 0 \) and \( V_{n+1}^{\ast} = \mathbb{G} V_n^{\ast} \). Then, we have \( V^{\ast} = \lim_{n \to \infty} V_n^{\ast} \) and \( V^{\ast} = \mathbb{G} V^{\ast} \).

**Proof** We define the policy \( f^{\ast} \) for the sMDPs in (4.1) by

\[
f^{\ast}(s, x) = \mathbb{1}_{(0, \infty) \times E}(s, x) \mathbb{1}_{\{U^{\ast}(s, x) = \hat{g}(x, 1)\}}(s, x) + \mathbb{1}_{\{\Delta\}}(x). \quad (4.18)
\]

Since \( \{ U^{\ast}(s, x) = g(x, 1) \} \) is a measurable subset of \( \mathbb{R} \times \hat{E} \) and \( f^{\ast}(\cdot, \Delta) = 1 \), \( f^{\ast} \) is a deterministic stationary policy. Moreover, for any \( x \in E \), we have that \( f^{\ast}(0, x) = 0 \), which implies \( f^{\ast} \in \Pi^0_{D_H} \). Next, we will show that \( f^{\ast} \) is \( T \)-optimal.

(i) \( x = \Delta \) : noting that \( A(\Delta) = \{ 1 \} \), by \( U_0^{\ast} \equiv 0 \) and \( \hat{c}(\Delta, 1) = \hat{g}(\Delta, 1) = 0 \), we have that \( U_n^{\ast}(s, \Delta) = 0 \) for all \( n \geq 0 \) and \( s \in [0, T] \), which implies that

\[
U^{\ast}(s, \Delta) = \mathbb{T} U^{\ast}(s, \Delta) = \mathbb{T} f^{\ast}(s, \Delta) U^{\ast}(s, \Delta) = 0 \quad \forall s \in [0, T].
\]
(ii) $s = 0$ and $x \in E$: by $f^*(0, x) = 0$, we have $U^*(0, x) = \min\{0, g(x)\} = \mathbb{T} f^*(0, x) U^*(0, x)$.

(iii) $(s, x) \in (0, T] \times E$: noting that $\mathbb{T}^1 U^*(s, x) = \hat{g}(x, 1)$, by $U^* = \mathbb{T} U^*$, we have

$$U^*(s, x) = \begin{cases} \mathbb{T}^1 U^*(s, x) = \mathbb{T} f^*(s, x) U^*(s, x), & U^*(s, x) = \hat{g}(x, 1); \\ \mathbb{T}^0 U^*(s, x) = \mathbb{T} f^*(s, x) U^*(s, x), & U^*(s, x) \neq \hat{g}(x, 1). \end{cases}$$

Hence, part (c) of Theorem 3.11 ensures that $f^*$ is $T$-optimal. Furthermore, Theorem 4.4 implies that

$$U^*(s, x) = U f^*(s, x) = V^{\tau_{f^*}}(s, x) \geq V^*(s, x) \quad \forall x \in E, \ s \in [0, T],$$

where $\tau_{f^*}$ is the stopping time induced by $f^*$ and $s$. On the other hand, for each $\tau \in \Gamma$, Theorem 4.7 gives

$$V^\tau(s, x) = U^{\tau \pi}(s, x) \geq U^*(s, x) \quad \forall x \in E, \ s \in [0, T].$$

By the arbitrariness of $\tau$, it holds that $V^*(s, x) \geq U^*(s, x)$. Hence,

$$V^*(s, x) = U^*(s, x) \quad \forall x \in E, \ s \in [0, T]. \quad (4.19)$$

Next, we show that for all $n \geq 0$

$$U^*_n(s, x) = V^*_n(s, x) \quad \forall x \in E, \ s \in [0, T]. \quad (4.20)$$

When $n = 0$, we have $U^*_0(s, x) = V^*_0(s, x) = 0$. Assume that (4.20) holds for some $n$. Then,

$$U^*_{n+1}(s, x) = \min \left\{ \hat{g}(x, 1), \hat{c}(x, 0) \int_0^s (1 - \hat{Q}_T(t, \hat{E}|x, 0)) dt \\
+ \hat{g}(x, 0)(1 - \hat{Q}_T(s, \hat{E}|x, 0)) \\
+ \int_{[0,s]} \int_{\hat{E}} U^*_n(s - t, y) \hat{Q}_T(dt, dy|x, 0) \right\}$$

$$= \min \left\{ g(x), c(x) \int_0^s (1 - Q(t, E|x)) dt \\
+ \int_{[0,s]} \int_E V^*_n(s - t, y) Q(dt, dy|x) \right\}$$

$$= \mathbb{G} V^*_n(s, x) = V^*_{n+1}(s, x),$$

where the second equality is due to inductive hypothesis and (4.2)-(4.4). Hence, by part (a) of Theorem 3.11, (4.19) and (4.20), we obtain $V^*(s, x) = \lim_{n \to \infty} V^*_n(s, x)$. Furthermore, the monotone convergence theorem ensures $V^* = \mathbb{G} V^*$.  

\[ \square \]
Theorem 4.9 \textit{(T-optimal stopping time.)} Suppose that Assumption 2.2 holds. Define a subset of $(0, T] \times E$

\[ D^* := \{(s, x) \in (0, T] \times E : V^*(s, x) = g(x)\}, \]

and for each $\omega = (x_0, t_1, x_1, \ldots, t_n, x_n, \ldots) \in \Omega$, define

\[ \tau^*(\omega) := \inf \left\{ n \in \mathbb{N} \mid (T - \sum_{k=1}^{n} t_k, x_n) \in D^* \right\} \wedge \inf \left\{ n \in \mathbb{N} \mid T \leq \sum_{k=1}^{n} t_k \right\} \quad (4.22) \]

Then, $\tau^*$ is a T-optimal stopping time and satisfies that $\mathbb{P}_x(\tau^* = \infty) = 0$ for all $x \in E$.

\textbf{Proof} The definition of $f^*$ given in (4.18) and Definition 4.2 show that for each $\omega \in \Omega$

\[ \tau_{f^*}^T(\omega) = \inf \left\{ n \in \mathbb{N} \mid f^*(T - S(\omega), X_n(\omega)) = 1 \right\} \]

\[ = \inf \left\{ n \in \mathbb{N} \mid U^*(T - \sum_{k=1}^{n} t_k, x_n) = g(x_n), T > \sum_{k=1}^{n} t_k \right\} \]

\[ = \inf \left\{ n \in \mathbb{N} \mid (T - \sum_{k=1}^{n} t_k, x_n) \in D^* \right\}, \]

which implies $\tau^*(\omega) = \tau_{f^*}^T(\omega) \wedge \inf \left\{ n \in \mathbb{N} \mid T \leq \sum_{k=1}^{n} t_k \right\}$. Note that $\tau_{f^*}^T$ may not be a finite stopping time, whereas, using (4.22), we have

\[ 0 \leq \mathbb{P}_x(\tau^* = \infty) \leq \mathbb{P}_x(T \geq \lim_{n \to \infty} S_n) = 0 \quad \forall x \in E, \]

which means that $\tau^*$ is a finite stopping time. Next, we will show that $\tau^*$ is T-optimal. Since $V^*(0, x) = 0 = V^+(0, x)$, we only need to consider the case $T > 0$. First, we have

\[ \{\tau^* = 0\} = \{\tau_{f^*}^T = 0\} \cup \{\tau_{f^*}^T > 0, T \leq 0\} = \{\tau_{f^*}^T = 0\}; \]

\[ \{\tau^* = n + 1\} = \{\tau_{f^*}^T = n + 1\} \cup \{\tau_{f^*}^T > n + 1, S_n < T \leq S_{n+1}\}, \quad \forall n \geq 0; \]

\[ \{\tau^* = \infty\} = \{\tau_{f^*}^T = \infty\} \cap \{T \geq \lim_{n \to \infty} S_n\}. \]

Hence, using Theorem 4.4 and Theorem 4.8, we obtain that

\[ V^+(T, x) = \sum_{n=1}^{\infty} \left\{ \mathbb{E}_x \left[ \mathbb{1}_{\{\tau_{f^*}^T = n\}} R_n^T \right] + \mathbb{E}_x \left[ \mathbb{1}_{\{S_{n-1} < T \leq S_n\}} \mathbb{1}_{\{\tau_{f^*}^T > n\}} R_n^T \right] \right\} \]

\[ + \mathbb{E}_x \left[ \mathbb{1}_{\{\tau_{f^*}^T = 0\}} R_n^T \right] \]
Given any \( n \geq 0 \), for any \( \varepsilon > 0 \), the cost of the stopping time \( \tau_\varepsilon \) is \( \varepsilon \)-\( T \)-optimal if it holds that \( V^{\tau}(T, x) - V^{\tau_\varepsilon}(T, x) \leq \varepsilon \) for all \( x \in E \), where \( V^{\tau}(T, x) \) is \( T \)-horizon expected cost of the stopping time \( \tau \) given in (2.4).

The following theorem shows that for any \( \varepsilon > 0 \), we can iterate enough times and get an \( \varepsilon \)-\( T \)-optimal stopping time under some conditions. For the convenience of statement, we give two notations, i.e., \( \| f \| := \sup_{x \in C} | f(x) | \) for any function \( f \) defined on the set \( C \); \( [ x ] := \min \{ n \in \mathbb{N} : n \geq x \} \) for any \( x \in \mathbb{R}_+ \).

**Theorem 4.11** Suppose that the semi-Markov kernel \( Q \) satisfies \( \sup_{x \in E} Q(T, E|x) =: \beta < 1 \) and \( T \| c \| + \| g \| =: M < \infty \). Given any \( n \geq 0 \), define \( D_n \) the subset of \( (0, T] \times E \) as

\[
D_n := \{ (s, x) \in (0, T] \times E \mid g(x) = \mathbb{G}V^*_n(s, x) \}.
\]

Moreover, for each \( \omega = (x_0, t_1, \ldots, x_n, t_{n+1}, \ldots) \in \Omega \), define

\[
\tau_n(\omega) = \inf \left\{ m \in \mathbb{N} \mid ((T - \sum_{k=1}^m t_k, x_m) \in D_n) \land \inf \left\{ m \in \mathbb{N} \mid T \leq \sum_{k=1}^m t_k \right\} \right\}.
\]

Then, the following statements hold.

(a) For any \( \varepsilon > 0 \), if \( \| V^*_n \| \leq \frac{1-\beta}{\beta} \varepsilon \) \( n \geq 0 \), \( \tau_N \) is an \( \varepsilon \)-\( T \)-optimal stopping time.

(b) Given any \( n \geq 0 \), it holds that \( V^* - V^*_n \| \leq \beta^n M. \) Furthermore, define the number of iterations \( n_\varepsilon \) as \( n_\varepsilon := \left[ \frac{\log(\varepsilon (1-\beta))/\log \beta}{\log(\varepsilon (1-\beta))/\log \beta + \log(M+1)} \right] \), then \( \tau_{n_\varepsilon} \) is an \( \varepsilon \)-\( T \)-optimal stopping time.

**Proof** (a) Note that \( \sup_{x \in E} Q(T, E|x) = \beta < 1 \) gives that Assumption 2.2 holds with \( \delta = T \) and \( \alpha = 1 - \beta \). Moreover, by (4.2), we have that

\[
\hat{Q}(T, \hat{E}|x, a) = \begin{cases} Q(T, E|x) \leq \beta, & x \in E, a = 0; \\
1_{[T+1, +\infty)}(T)\delta_\Delta(\hat{E}) = 0 \leq \beta, & x \in \hat{E}, a = 1.
\end{cases}
\]
Define a policy $f_N$ by

$$f_N(s, x) = 1_{(0, \infty)}(s) \mathbb{I}_{\{TU_N^*(s, x) = \tilde{g}(x, 1)\}}(s, x) + \mathbb{I}_\Delta(x).$$

and then, we have $f_N \in \Pi^{0}_{HD}$. Moreover, similar to the proof of Theorem 4.8, we have

$$U_N^*(s, x) = \mathbb{T} U_N^*(s, x) = \mathbb{T} f_N(s, x) U_N^*(s, x), \quad \forall (s, x) \in [0, T] \times \hat{E}. \quad (4.23)$$

Next, we use the induction to show the following inequality:

$$U_N^*(s, x) \geq U_N^*(s, x) + \hat{\beta} f_N(s, x) \left[ U_N^*(s - \tilde{S}_{n+1}^+, \tilde{X}_{n+1}) \right] - \beta \|U_{N+1}^* - U_N^*\|. \quad (4.24)$$

By the definition of $\mathbb{T} f_N(s, x)$ and (4.23), we obtain

$$U_N^*(s, x) = U_0^*(s, x) + \int_{[0, x]} \int_{E} U_N^*(s - t, y) \hat{Q}_T(dt, dy|x, f_N(s, x))$$

$$\geq U_0^*(s, x) + \hat{\beta} f_N(s, x) \left[ U_N^*(s - \tilde{S}_1^+, \tilde{X}_1) \right] - \beta \|U_{N+1}^* - U_N^*\|,$n

which means that (4.24) holds for $n = 0$. For any $n \geq 0$, it holds that

$$\hat{\beta} f_N(s, x) \left[ U_N^*(s - \tilde{S}_n^{+, \infty}) \right]$$

$$= \int_{\hat{E}} \delta_x(dx_0) \int_{[0, x]} \int_{E} \hat{Q}_T(dt_1, dx_1|x_0, f_N(s, x_0))$$

$$\int_{[0, s_{s, 1}]} \int_{E} \hat{Q}_T(dt_2, dx_2|x_1, f_N(s - s_1, x_1))$$

$$\cdots \int_{[0, s_{s, n}]} \int_{E} \hat{Q}_T(dt_{n+1}, dx_{n+1}|x_n, f_N(s - s_{n}, x_{n}))(U_N^*(s - s_{n, 1}, x_{n+1})$$

$$\geq \int_{\hat{E}} \delta_x(dx_0) \int_{0}^{\infty} \int_{E} \hat{Q}_T(dt_1, dx_1|x_0, f_N(s, x_0))$$

$$\int_{[0, s_{s, 1}]} \int_{E} \hat{Q}_T(dt_2, dx_2|x_1, f_N(s - s_1, x_1))$$

$$\cdots \int_{[0, s_{s, n}]} \int_{E} \hat{Q}_T(dt_{n+1}, dx_{n+1}|x_n, f_N(s - s_{n}, x_{n}))(U_N^*(s - s_{n, 1}, x_{n+1})$$

$$- \|U_{N+1}^* - U_N^*\|$$

$$\geq \hat{\beta} f_N(s, x) \left[ U_N^*(s - \tilde{S}_{n+1}^+, \tilde{X}_{n+1}) \right] - \beta \|U_{N+1}^* - U_N^*\|. \quad \square$$
Thus, suppose that (4.24) holds for some $n \geq 0$, by (3.3), (3.4), and (4.23), we have

$$U_{N+1}^n(s, x) \geq U_{n+1}^f(s, x) + \frac{\beta}{1-\beta} \left[ T_{N+1}' U_{N+1}^*((s - \hat{S}_{n+1})^+, \hat{X}_{n+1}) \right]$$

$$- \sum_{k=0}^{n} B^{k+1} ||U_{n+1}^* - U_N^*||$$

$$= U_{n+1}^f(s, x) + \frac{\beta}{1-\beta} \left[ U_N^*((s - \hat{S}_{n+2})^+, \hat{X}_{n+2}) \right]$$

$$- \sum_{k=0}^{n} B^{k+1} ||U_{n+1}^* - U_N^*||$$

$$\geq U_{n+1}^f(s, x) + \frac{\beta}{1-\beta} \left[ U_{n+1}^*((s - \hat{S}_{n+2})^+, \hat{X}_{n+2}) \right]$$

$$- \sum_{k=0}^{n+1} B^{k+1} ||U_{n+1}^* - U_N^*||.$$ 

Therefore, we obtain (4.24). Then, passing the limit $n \to \infty$ in (4.24), it holds that for all $(s, x) \in [0, T] \times \hat{E}$

$$U^*(s, x) \geq U_{N+1}^*(s, x) \geq U_N^f(s, x) - \frac{\beta}{1-\beta} ||U_{n+1}^* - U_N^*|| \quad \forall n \geq 0. \quad (4.25)$$

On the other hand, recalling that $U_{N+1}^*(s, \Delta) = U_N^*(s, \Delta) = 0$ for all $s \in [0, T]$, by (4.20), we have

$$||U_{N+1}^* - U_N^*|| = \sup_{(s, x) \in [0, T] \times \hat{E}} ||U_{N+1}^*(s, x) - U_N^*(s, x)||$$

$$= \sup_{(s, x) \in [0, T] \times \hat{E}} ||V_{N+1}^*(s, x) - V_N^*(s, x)|| \leq \frac{1-\beta}{\beta} \varepsilon.$$ 

Therefore, it holds that

$$U^*(s, x) \geq U_N^f(s, x) - \varepsilon \quad \forall (s, x) \in [0, T] \times \hat{E}.$$ 

In the same method of Theorem 4.9, we can verify $\tau_N = T_{N}^f \wedge \inf\{n \in \mathbb{N} | T \leq S_n\}$ and

$$V_{\tau_N}(T, x) = U_N^f(T, x) \leq V^*(T, x) + \varepsilon \quad \forall x \in \hat{E},$$

i.e., $\tau_N$ is an $\varepsilon$-$T$-optimal stopping time.

(b) By Theorem 4.8, for any $n \geq 1$ and $(s, x) \in [0, T] \times \hat{E}$, we have

$$|V^*(s, x) - V_n^*(s, x)| \leq \int_E \int_{[0, x]} V^*(s - t, y) Q(dt, y|x)$$

$$\square$$ Springer
\[-\int_E \int_{[0,s]} V_{n-1}^*(s-t, y) Q(dt, dy) \leq \beta \| V^* - V_{n-1}^* \|,\]

which implies that
\[\| V^* - V_n^* \| \leq \beta \| V^* - V_{n-1}^* \| \leq \beta^n \| V^* - V_0^* \| \leq \beta^n (T \| c \| + \| g \|) = \beta^n M.\]

Moreover, noting that \( V_n^* \leq V_{n+1}^* \leq V^* \), the definition of \( n_\varepsilon \) gives that
\[\| V_{n_\varepsilon+1}^* - V_n^* \| \leq \| V^* - V_{n_\varepsilon}^* \| \leq \beta^{n_\varepsilon} M \leq \frac{1 - \beta}{\beta} \varepsilon,
\]

which, together with part (a), implies that \( \tau_{n_\varepsilon} \) is an \( \varepsilon\)-\( T \)-optimal stopping time. \( \square \)

From Theorem 4.11, we can derive an algorithm for computing \( \varepsilon\)-\( T \)-optimal stopping times.

**Value Iteration Algorithm.** Here, we introduce the algorithm for the \( \varepsilon\)-\( T \)-optimal stopping times.

**Step 1 (Initialization.)** Let \( V_0^*(s, x) = 0 \) for every \((s, x) \in [0, T] \times E\).

**Step 2 (Iteration.)** Compute the function \( V_n^*(s, x) \) for every \((s, x) \in [0, T] \times E\) by
\[
V_{n+1}^*(s, x) = \min \left\{ c(x) \int_0^s (1 - Q(t, E|x)) dt + \int_E \int_{[0,s]} V_n^*(s-t, y) Q(dt, dy|x), g(x) \right\}.
\]

**Step 3 (Accuracy control.)** If \( V_{n+1}^*(s, x) - V_n^*(s, x) \leq \frac{1 - \beta}{\beta} \varepsilon \) for every \((s, x) \in [0, T] \times E\), go to Step 4; otherwise, go to Step 2 by replacing \( n \) with \( n + 1 \).

**Step 4 (\( \varepsilon\)-\( T \)-optimal stopping time.)** Compute the set \( D^\varepsilon = \left\{ (s, x) \in (0, T] \times E \mid g(x) = G V_n^*(s, x) \right\} \) and the \( \varepsilon\)-\( T \)-optimal stopping time
\[
\tau^\varepsilon(\omega) = \inf \left\{ m \in \mathbb{N} \mid ((T - \sum_{k=1}^m t_k, x_n) \in D^\varepsilon) \wedge \inf \left\{ m \in \mathbb{N} \mid T \leq \sum_{k=1}^m t_k \right\} \forall \omega \in \Omega. \right\}
\]

**5 Application**

In this section, we will apply our results to a house rental problem and show how to use the algorithm to compute the \( \varepsilon\)-\( T \)-optimal stopping time.
**Example 5.1** Suppose that a decision-maker has a batch of products and plans to sell it to a company (denoted by Company A) in \( T \) months (for any given \( T > 0 \)). If the decision-maker sells the products to other companies in advance, he/she needs to pay a penalty \( b \) to Company A. To store the products, the decision-maker needs to rent a warehouse from a landlord for \( T \) months. The rental rates for the warehouse depend on the supply–demand relationship of housing rental market, which usually are described as three states of “Surplus,” “Equilibrium,” and “Shortage.” Given any state \( x \), suppose that the sojourn time at \( x \) follows an exponential distribution with a parameter \( \mu(x) > 0 \) and the transition probability from \( x \) to a state \( y \) is denoted by \( p_{xy} \). When the supply–demand relationship changes before \( T \), the decision-maker needs to decide whether to continue renting the warehouse. If the decision-maker rents the warehouse, he/she pays the rent at rate \( c(x) \) to the landlord at state \( x \). However, if the decision-maker stops to rent the warehouse, which means he/she needs to sell the product to other companies in advance, and thus he/she needs to pay a penalty \( b \) to Company A as well as another penalty \( g(x) \) to the landlord.

For convenience, denote by 1 the state “Surplus,” by 2 the state “Equilibrium,” and by 3 the state “Shortage.” Naturally, the meaning of these states implies \( 0 \leq c(1) \leq c(2) \leq c(3) \), and \( g(1) \geq g(2) \geq g(3) \geq 0 \). The decision-maker wishes to find an optimal stopping time to minimize the cost.

We now formula this problem as an optimal stopping time problem with \( T \) finite horizon. Let \( E := \{1, 2, 3\} \) be the state space and \( Q(t, y|x) := (1 - e^{-\mu(x) t}) p_{xy} \) be the semi-Markov kernel. It can be verified that Assumption 2.2 holds for \( \delta = T \) and \( \alpha = \min\{e^{-T \mu(1)}, e^{-T \mu(2)}, e^{-T \mu(3)}\} \in (0, 1) \), which implies the existence of \( T \)-optimal stopping times (by Theorem 4.9). Moreover, since \( \sup_{x \in E} Q(T, E|x) = 1 - \min\{e^{-T \mu(1)}, e^{-T \mu(2)}, e^{-T \mu(3)}\} \), the conditions in Theorem 4.11 are satisfied. Therefore, for \( \varepsilon > 0 \), we can obtain an \( \varepsilon \)-\( T \)-optimal stopping time.

Next, we show the effectiveness of the algorithm in Sect. 4 through a specific set of data. Suppose that \( T = 20 \), the parameters \( \mu(\cdot) \) are \( \mu(1) = 0.4, \mu(2) = 0.2, \) and \( \mu(3) = 0.3 \), the transition probability matrix is given as

\[
p_{xy} = \begin{pmatrix} 0.35 & 0.65 & 0 \\ 0.2 & 0.6 & 0.2 \\ 0 & 0.7 & 0.3 \end{pmatrix},
\]

the penalty to Company A is \( b = 800 \) and

\[
c(1) = 95, \ c(2) = 135, \ c(3) = 180; \quad g(1) = 700, \ g(2) = 300, \ g(3) = 100.
\]

Take \( \varepsilon := 10^{-8} \). Using the data and the algorithm above, we compute \( V_n^*(s, x) \) in MATLAB and obtained \( \|V_{23}^* - V_{22}^*\| \leq 10^{-14} \leq \frac{1-\beta}{\beta} \varepsilon \), where \( \beta = 1 - e^{-8} \). Thus, by Theorem 4.11, we have \( \|V_{23}^* - V^*\| \leq 10^{-8} \) and \( \tau_{22} \) is the \( \varepsilon \)-\( T \)-optimal stopping time given (5.1).

Figure 1 shows the functions \( V_n^*(\cdot, x) \) for each \( x \in E \). From Fig. 1, we obtain the set
Fig. 1 The functions $V_{23}^\ast(\cdot, x)$

$$D_{22} = \{(s, x) \in (0, T) \mid g(x) = \mathbb{G}V_{22}^\ast(s, x) = V_{23}^\ast(s, x)\}$$

$$= \{(s, 2) \mid s \in [8.31, 20]\} \cup \{(s, 3) \mid s \in [5.97, 20]\},$$

and the $\varepsilon$-$T$-optimal stopping time $\tau_{22}$ is given as

$$\tau_{22}(\omega) = \inf \left\{ m \in \mathbb{N} \mid (T - \sum_{k=1}^{m} t_k, x_m) \in D_{22} \right\} \wedge \inf \left\{ m \in \mathbb{N} \mid T \leq \sum_{k=1}^{m} t_k \right\}$$

$$= \inf \left\{ n \in \mathbb{N} \mid x_n = 2 \text{ and } \sum_{k=1}^{n} t_k < 11.69, \text{ or } x_n = 3 \text{ and } \sum_{k=1}^{n} t_k < 14.03 \right\} \wedge \inf \left\{ n \in \mathbb{N} \mid \sum_{k=1}^{n} t_k \geq 20 \right\}. \quad (5.1)$$

for each $\omega = (x_0, t_1, \ldots, x_n, t_{n+1}, \ldots) \in \Omega$.

6 Conclusions

This paper deals with the optimal stopping problems on sMPs with finite horizon. The main contributions are as follows: 1) We propose an explicit construction method, which converts a given optimal stopping problem on sMPs with finite horizon to another equivalent special class of sMDPs. This equivalence is embodied in finite horizon expected costs of sMPs and a special class of sMDPs, which is explicitly constructed based on the data of the optimal stopping problems on sMPs. More specifically, every stopping time of sMPs can induce a policy of the constructed sMDPs such that
the expected costs are equal, and vice versa; see Theorem 4.7. 2) Motivating by the optimal stopping problems with terminal costs, we have extended the existing results on sMDPs to the case with additional terminal costs, establish the existence of an optimal policy, and give an iterative algorithm; see Theorem 3.11. 3) By the results on sMDPs developed here and the equivalence above, we not only give the existence of optimal stopping times on sMPS, but also provide a value iteration algorithm for computing them. By estimating the convergence rate of the algorithm, we prove that $\varepsilon(> 0)$-optimal stopping times can be obtained with a finite number of iterations. Finally, an example is shown to illustrate how to use our algorithm, see Example 5.1.

Note that using the equivalence of the optimal stopping problems on sMPS and sMDPs has the advantage of providing an algorithm, but at the same time it leads us to only consider the stopping times taking values in integers. Therefore, it is desirable and unsolved to give a similar equivalence of sMDPs and optimal stopping problems with real-valued stopping times in [5, 26] or randomized stopping times in [4, 15, 27], which can be used to derive an algorithm for computing corresponding $\varepsilon$-optimal stopping times.

Acknowledgements This work was partly supported by the National Natural Science Foundation of China (No. 11931018, 11701588). We would like to thank the referee for helpful comments and suggestions which led to the improved version of the paper.

References

1. Arkin, V.I., Slastnikov, A.D.: On optimal threshold stopping times for Itô diffusions. Stochastics. 93(5), 655–681 (2021) https://doi.org/10.1080/17442508.2020.1783263
2. Bäuerle, N., Popp, A.: Risk-sensitive stopping problems for continuous-time Markov chains. Stochastics 90(3), 411–431 (2018). https://doi.org/10.1080/17442508.2017.1357724
3. Bäuerle, N., Rieder, U.: Markov Decision Processes with Applications to Finance. Springer, Heidelberg (2011)
4. Belomestny, D., Krätschmer, V.: Optimal stopping under model uncertainty: randomized stopping times approach. Ann. Appl. Probab. 26(2), 1260–1295 (2016). https://doi.org/10.1214/15-AAP1116
5. Boshuizen, F.A., Gouweleeuw, J.M.: General optimal stopping theorems for semi-Markov processes. Adv. Appl. Probab. 25(4), 825–846 (1993). https://doi.org/10.1017/S0001867800025775
6. Brémaud, P.: Point Processes and Queues: Martingale Dynamics. Springer, New York (1981)
7. Çekyay, B.: Customizing exponential semi-Markov decision processes under the discounted cost criterion. Eur. J. Oper. Res. 266(1), 168–178 (2017). https://doi.org/10.1016/j.ejor.2017.09.016
8. Chen, F., Guo, X.P., Liao, Z.-W.: Optimal stopping time on discounted semi-Markov processes. Front. Math. China. 16(2), 303–324 (2021). https://doi.org/10.1007/s11464-021-0919-4
9. Chow, Y.S., Robbins, H., Siegmund, D.: Great Expectations: The Theory of Optimal Stopping. Houghton Mifflin Company, Boston (1991)
10. Christensen, S., Lindensjö, K.: On time-inconsistent stopping problems and mixed strategy stopping times. Stochastic Process. Appl. 130(5), 2886–2917 (2020). https://doi.org/10.1016/j.spa.2019.08.010
11. Davis, M.H.A.: Markov Models and Optimization. Chapman and Hall, London (1993)
12. Dufour, F., Piunovskiy, A.B.: Multiobjective stopping problem for discrete-time Markov processes: convex analytic approach. J. Appl. Probab. 47(4), 947–966 (2010). https://doi.org/10.1017/S0021900200007282
13. Feinberg, E.A.: Reduction of discounted continuous-time MDPs with unbounded jump and reward rates to discrete-time total-reward MDPs. Optimization, Control, and Applications of Stochastic Systems, Systems Control Found. Birkhuser, New York, 77–97 (2012) https://doi.org/10.1007/978-0-8176-8337-5_5
14. Gapeev, P.V., Kort, P.M., Lavrutich, M.N.: Discounted optimal stopping problems for maxima of geometric Brownian motions with switching payoffs. Adv. Appl. Probab. 53(1), 189–219 (2021). https://doi.org/10.1017/apr.2020.57
15. Henderson, V., Hobson, D., Zeng, M.: Optimal stopping and the sufficiency of randomized threshold strategies. Electron. Commun. Probab. 23(30), 11 (2018). https://doi.org/10.1214/18-ECP125
16. Hernández-Lerma, O., Lasserre, J.B.: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer, New York (1996)
17. Huang, Y.H., Guo, X.P.: Finite horizon semi-Markov decision processes with application to maintenance systems. Eur. J. Oper. Res. 212(1), 131–140 (2011). https://doi.org/10.1016/j.ejor.2011.01.027
18. Huang, Y.J., Zhou, Z.: The optimal equilibrium for time-inconsistent stopping problems: the discrete-time case. SIAM J. Control. Optim. 57(1), 590–609 (2019). https://doi.org/10.1137/17M1139187
19. Jaśkiewicz, A., Nowak, A.S.: Optimality in Feller semi-Markov control processes. Oper. Res. Lett. 34(6), 713–718 (2006). https://doi.org/10.1016/j.orl.2005.11.005
20. Karpowicz, A.: Double optimal stopping in the fishing problem. J. Appl. Probab. 46(2), 415–428 (2009). https://doi.org/10.1239/jap/1245676097
21. Karpowicz, A., Szajowski, K.: Double optimal stopping of a risk process. Stochastics 79(1–2), 155–167 (2007). https://doi.org/10.1080/17442500601084204
22. Kitaev, M.Y.: Semi-Markov and jump Markov controlled models: average cost criterion. Theory Probab. Appl. 30(2), 272–288 (1986). https://doi.org/10.1137/1130036
23. Kitaev, M.Y., Rykov, V.: Controlled Queueing Systems. CRC Press, Boca Raton (1995)
24. Linnios, N., Oprisan, G.: Semi-Markov Processes and Reliability. Birkhäuser, Boston (2001)
25. Nutz, M., Zhang, Y.: Conditional optimal stopping: A time-inconsistent optimization. Ann. Appl. Probab. 30(4), 1669–1692 (2020). https://doi.org/10.2139/ssrn.3409585
26. Ohtsubo, Y.: Optimal stopping in generalized semi-Markov jump processes. Mem. Fac. Sci. A 5, 63–71 (1984)
27. Rosenberg, D., Solan, E., Vieille, N.: Stopping games with randomized strategies. Probab. Theory Related Fields 119(3), 433–451 (2001). https://doi.org/10.1007/PL00008766
28. de Saporta, B., Dufour, F., Nivot, C.: Partially observed optimal stopping problem for discrete-time Markov processes. 4OR. 15(3), 277–302 (2017) https://doi.org/10.1007/s10288-016-0337-8
29. Shaï, J.H., Tian, T.R.: Optimal stopping problem for jump-diffusion processes with regime-switching. Nonlinear Anal. Hybrid Syst. 41: No.101029 (2021) https://doi.org/10.1016/j.nahs.2021.101029
30. Shiryaev, A.N.: Optimal Stopping Rules. Springer, New York (1978)
31. Sinha, S., Mondal, P.: Semi-Markov decision processes with limiting ratio average rewards. J. Math. Anal. Appl. 455(1), 864–871 (2017). https://doi.org/10.1016/j.jmaa.2017.06.017
32. Szajowski, K.: Optimal stopping of a 2-vector risk process. Banach Center Publ. 90, 179-191 (2010) https://doi.org/10.4064/bc90-0-12
33. Zuckerman, D.: Optimal stopping in a semi-Markov shock model. J. Appl. Probab. 15(3), 629–634 (1978). https://doi.org/10.2307/3213126

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