Chapter 11
Constrained Optimality for First Passage Criteria in Semi-Markov Decision Processes

Yonghui Huang and Xianping Guo

11.1 Introduction

In the field of Markov decision problems (MDPs), the control horizon is usually a fixed finite interval $[0, T]$ or the infinite interval $[0, +\infty)$. In many real applications, however, the control horizon may be a *random duration* $[0, \tau]$, where the terminal time τ is a random variable at which the state of the controlled system changes critically and the control beyond τ may no longer be meaningful or necessary. For example, in the insurance systems [27], the control after the time when the company is bankrupt becomes unnecessary. Therefore, it makes better sense to consider the problem in $[0, \tau]$, where τ represents the bankruptcy time of the company. Such situations motivate *first passage* problems in MDPs [13, 15, 19, 21, 22, 28], for which one generally aims at maximizing/minimizing the expected reward/cost over a first passage time to some target set.

This chapter is devoted to studying constrained optimality for first passage criteria, for which the dynamic of a system is described by semi-Markov decision processes (SMDPs). The state space is assumed to be denumerable, while the action set is general. Both reward and cost rates are possibly *unbounded*. A key feature of our model is that the discount rate is state-action dependent, and furthermore, the undiscounted case is allowed. This feature makes our model more general since the state-action-dependent discount rate exactly characterizes the practical cases such as the interest rate in economic and financial systems [2, 9, 17, 23, 26], which can be adjusted according to the real circumstances. We aim to maximize the expected reward obtained during a first passage time to some target set, subject to that the associated expected cost over this first passage time does not exceed a given constraint. An interesting special case is that in which the reward rates are uniformly

Y. Huang • X. Guo (✉)
Sun Yat-Sen University, Guangzhou 510275, China
e-mail: hyongh5@mail.sysu.edu.cn; mcsgxp@mail.sysu.edu.cn

D. Hernández-Hernández and A. Minjárez-Sosa (eds.), *Optimization, Control, and Applications of Stochastic Systems*, Systems & Control: Foundations & Applications,
DOI 10.1007/978-0-8176-8337-5_11,

equal to one, which corresponds to a stochastic time optimal control problem with a target set; see Remark 11.2.4(d) for details.

Previously, Beutler and Ross [3] consider constrained SMDPs with the *long-run average* criteria. They suppose that the state space of the SMDP is finite, and the action space compact metric. A Lagrange multiplier formulation involving a dynamic programming equation is utilized to relate the constrained optimization to an unconstrained optimization parametrized by the multiplier. This approach leads to a proof for the existence of a semi-simple optimal constrained policy. That is, there is at most one state for which the action is randomized between two possibilities; at all other states, an action is uniquely chosen for each state. Feinberg [4] further investigates *constrained average reward* SMDPs with finite state and action sets. They develop a technique of state-action renewal intensities and provide linear programming algorithms for the computation of optimal policies. On the other hand, Feinberg [5] deals with constrained *infinite horizon discounted* SMDPs. Compared with the existing works above, however, our main interest in this chapter is to analyze the constrained optimality for *first passage* criteria in SMDPs, which, to best of our knowledge, is an issue not yet explored.

To obtain the existence of a constrained *first passage* optimal policy, we first give suitable conditions and then employ the so-called Lagrange multiplier technique to analyze the constrained control problem. Based on the Lagrange multiplier technique, we transform the constrained control problem to an unconstrained one, prove that a constrained optimal policy exists, and show that the constrained optimal policy randomizes between two stationary policies differing in at most one state.

The rest of this chapter is organized as follows. In Sect. 11.2, we formulate the control model, followed by the optimality conditions and the main results on the existence of constrained optimal policies. In Sect. 11.3, some technique preliminaries are given, and the proof of the main result is presented in Sect. 11.4.

11.2 The Control Model

The model of constrained SMDPs considered in this chapter is specified by the eight objects

$$\{E, B, (A(i) \subset A, i \in E), Q(\cdot, \cdot \mid i,a), r(i,a), c(i,a), \alpha(i,a), \gamma\}, \tag{11.1}$$

where E is the *state space*, a denumerable set; $B \subset E$ is the given *target set*, such as the set of all bad states or of good states of a system; A is the *action space*, a Borel space endowed with the Borel σ-field $\mathscr{A}$; and $A(i) \in \mathscr{A}$ is the set of *admissible actions* at state $i \in E$. The transition mechanism of the SMDPs is defined by the *semi-Markov kernel* $Q(\cdot, \cdot|i,a)$ on $R_+ \times E$ given K, where $R_+ = [0, +\infty)$, and $K = \{(i,a) \mid i \in E, a \in A(i)\}$ is the set of feasible state-action pairs. It is assumed that (1) $Q(\cdot, j|i,a)$ (for any fixed $j \in E$ and $(i,a) \in K$) is a nondecreasing, right continuous real function on R_+ such that $Q(0, j|i,a) = 0$; (2) $Q(t, \cdot|\cdot, \cdot)$ (for each fixed $t \in R_+$) is a sub-stochastic kernel on E given K; and (3) $P(\cdot|\cdot, \cdot) := Q(\infty, \cdot|\cdot, \cdot)$ is a stochastic kernel on E given K. If action $a \in A(i)$ is selected in state i, then $Q(t, j \mid i,a)$ is the

joint probability that the sojourn time in state i is not greater than $t \in R_+$, and the next state is j. Moreover, $r(i,a)$ and $c(i,a)$ in (11.1) denote the *reward* and *cost rate functions* on K valued in $R = (-\infty, +\infty)$, respectively, which are both assumed to be measurable on $A(i)$ for each fixed $i \in E$. In addition, $\alpha(i,a)$ represents the *discount rate*, which is a measurable function from K to R_+. Finally, γ is a given *constraint constant*.

Remark 11.2.1. Compared with the models of the standard constrained discounted and average criteria [3–5], in this model (11.1), we introduce a target set $B \subset E$ of the controlled system, and furthermore, the discount rate $\alpha(i,a)$ here is state-action dependent and may be equal to zero (i.e., the undiscounted case is allowed).

To state the constrained SMDPs we are concerned with, we need to introduce the classes of policies. For each $n \geq 0$, let H_n be the family of admissible histories up to the nth jump (decision epoch), that is, $H_n = (R_+ \times K)^n \times (R_+ \times E)$, for $n = 0, 1, \ldots$.

Definition 11.2.1. A randomized history-dependent policy, or simply a policy, is a sequence $\pi = \{\pi_n, n \geq 0\}$ of stochastic kernels π_n on A given H_n satisfying

$$\pi_n(A(i_n) \mid h_n) = 1 \quad \forall\, h_n = (t_0, i_0, a_0, \ldots, t_{n-1}, i_{n-1}, a_{n-1}, t_n, i_n) \in H_n, \quad n = 0, 1, \ldots.$$

The class of all policies is denoted by Π. To distinguish the subclasses of Π, we let Φ be the family of all stochastic kernels φ on A given E such that $\varphi(A(i) \mid i) = 1$ for all $i \in E$, and $\mathbb{F}$ the set of all functions $f : E \to A$ such that $f(i)$ is in $A(i)$ for every $i \in E$. A policy $\pi = \{\pi_n\} \in \Pi$ is said to be *randomized Markov* if there is a sequence $\{\varphi_n\}$ of $\varphi_n \in \Phi$ such that $\pi_n(\cdot \mid h_n) = \varphi_n(\cdot \mid i_n)$ for every $h_n \in H_n$ and $n \geq 0$. We denote such a policy by $\pi = \{\varphi_n\}$. A randomized Markov policy $\pi = \{\varphi_n\}$ is said to be *randomized stationary* if every φ_n is independent of n. In this case, we write $\pi = \{\varphi, \varphi, \ldots\}$ as φ for simplicity. Further, a randomized Markov policy $\pi = \{\varphi_n\}$ is said to be *deterministic Markov* if there is a sequence $\{f_n\}$ of $f_n \in \mathbb{F}$ such that $\varphi_n(\cdot \mid i)$ is the Dirac measure at $f_n(i)$ for all $i \in E$ and $n \geq 0$. We write such a policy as $\pi = \{f_n\}$. In particular, a deterministic Markov policy $\pi = \{f_n\}$ is said to be (deterministic) *stationary* if f_n are all independent of n. Similarly, we write $\pi = \{f, f, \ldots\}$ as f for simplicity. We denote by Π_{RM}, $\Pi_{\mathrm{RS}}, \Pi_{\mathrm{DM}}$, and Π_{DS} the families of all randomized Markov, randomized stationary, deterministic Markov, and stationary policies, respectively. Obviously, $\Phi = \Pi_{\mathrm{RS}} \subset \Pi_{\mathrm{RM}} \subset \Pi$ and $\mathbb{F} = \Pi_{\mathrm{DS}} \subset \Pi_{\mathrm{DM}} \subset \Pi$.

Let $\mathbb{P}(E)$ denote the set of all the probability measures on E. For each $(s,\mu) \in R_+ \times \mathbb{P}(E)$ and $\pi \in \Pi$, by the well-known Tulcea's theorem ([10, Proposition C.10]), there exist a unique probability space $(\Omega, \mathscr{F}, P^{\pi}_{(s,\mu)})$ and a stochastic process $\{T_n, J_n, A_n, n \geq 0\}$ such that, for each $i, j \in E, t \in R_+, C \in \mathscr{A}$ and $n \geq 0$,

$$P^{\pi}_{(s,\mu)}(T_0 = s, J_0 = i) = \mu(i), \tag{11.2}$$

$$P^{\pi}_{(s,\mu)}(A_n \in C \mid h_n) = \pi_n(C \mid h_n), \tag{11.3}$$

$$P^{\pi}_{(s,\mu)}(T_{n+1} - T_n \leq t, J_{n+1} = j \mid h_n, a_n) = Q(t, j \mid i_n, a_n), \tag{11.4}$$

where T_n, J_n, and A_n denote the nth decision epoch, the state, and the action chosen at the nth decision epoch, respectively. The expectation operator with respect to $P^{\pi}_{(s,\mu)}$ is denoted by $E^{\pi}_{(s,\mu)}$. In particular, if μ is the Dirac measure $\delta_i(\cdot)$ concentrated at some state $i \in E$, we write $P^{\pi}_{(s,\mu)}$ and $E^{\pi}_{(s,\mu)}$ as $P^{\pi}_{(s,i)}$ and $E^{\pi}_{(s,i)}$, respectively. For simplicity, $P^{\pi}_{(0,\mu)}$ and $E^{\pi}_{(0,\mu)}$ is denoted by P^{π}_{μ} and E^{π}_{μ}, respectively. Without loss of generality, in the following, we always set the initial decision epoch $T_0 = 0$ and omit it.

Remark 11.2.2. (a) The construction of the probability measure space $(\Omega, \mathscr{F}, P^{\pi}_{(s,\mu)})$ and the above properties (11.2)–(11.4) follow from those in Limnios and Oprisan [18, p.33] and Puterman [24, p.534–535].

(b) Let $X_0 := 0$, $X_n := T_n - T_{n-1}$ $(n \geq 0)$ denote the sojourn times between decision epochs (jumps). Then, the stochastic process $\{T_n, J_n, A_n, n \geq 0\}$ may be rewritten as the one $\{X_n, J_n, A_n, n \geq 0\}$.

To avoid the possibility of an infinite number of decision epochs within finite time, we make the following assumption that the system does not have accumulation points.

Assumption 11.2.1 *For all* $\mu \in \mathbb{P}(E)$ *and* $\pi \in \Pi$, $P^{\pi}_{\mu}(\{\lim_{n\to\infty} T_n = \infty\}) = 1$.

To verify Assumption 11.2.1, we can use a sufficient condition below.

Condition 11.2.2 *There exist constants* $\delta > 0$ *and* $\varepsilon > 0$ *such that*

$$Q(\delta, E \mid i,a) \leq 1 - \varepsilon \quad \forall (i,a) \in K.$$

Remark 11.2.3. In fact, Condition 11.2.2 is the standard regular condition widely used in SMDPs [5, 16, 20, 24, 25], which exactly implies Assumption 11.2.1 above.

Under Assumption 11.2.1, we define an underlying continuous-time state-action process $\{Z(t), W(t), t \in R_+\}$ corresponding to the stochastic process $\{T_n, J_n, A_n\}$ by

$$Z(t) = J_n, \quad W(t) = A_n, \ \text{ for } \ T_n \leq t < T_{n+1}, \quad t \in R_+ \text{ and } n \geq 0.$$

Definition 11.2.2. The stochastic process $\{Z(t), W(t)\}$ is called a (continuous-time) SMDP.

For the target set $B \subset E$, we consider the random variable

$$\tau_B := \inf\{t \geq 0 \mid Z(t) \in B\} \ \ (\text{with } \inf\emptyset := \infty),$$

which is the first passage time into the set B of the process $\{Z(t), t \in R_+\}$. Now, fix an initial distribution $\mu \in \mathbb{P}(E)$. For each $\pi \in \Pi$, the expected first passage reward and cost criteria are defined as follows:

$$V_r(\mu,\pi) := E^{\pi}_{\mu}\left[\int_0^{\tau_B} \mathrm{e}^{-\int_0^t \alpha(Z(u),W(u))\mathrm{d}u} r(Z(t),W(t))\mathrm{d}t\right], \tag{11.5}$$

$$V_c(\mu,\pi) := E^{\pi}_{\mu}\left[\int_0^{\tau_B} \mathrm{e}^{-\int_0^t \alpha(Z(u),W(u))\mathrm{d}u} c(Z(t),W(t))\mathrm{d}t\right]. \tag{11.6}$$

To introduce the constrained problem, for the constraint constant γ in (11.1), let

$$U := \{\pi \in \Pi \mid V_c(\mu,\pi) \le \gamma\}$$

be the set of "constrained" policies. We assume that $U \neq \emptyset$ throughout the following. Then, the optimization problem we are interested in is to maximize the expected first passage reward $V_r(\mu,\pi)$ over the set U, and our goal is to find a *constrained optimal* policy as defined below.

Definition 11.2.3. A policy $\pi^* \in U$ is called constrained optimal if

$$V_r(\mu,\pi^*) = \sup_{\pi\in U} V_r(\mu,\pi).$$

Remark 11.2.4. (a) It is worthwhile to point out that the expected first passage reward criterion $V_r(\mu,\pi)$ defined in (11.5) is different from the usual discounted reward criteria [11, 12, 24] and the total reward criteria without discount [6, 11, 24]. In fact, the former concerns with the performance of the system during a first passage time to some target set, while the latter concern with the performance of the system over an infinite horizon. However, if the target set $B = \emptyset$ (and thus $\tau_B \equiv \infty$) and, furthermore, the discount factor $\alpha(i,a)$ is state-action independent (say, $\alpha(i,a) \equiv \alpha$), then the expected first passage reward criterion (11.5) above will be directly reduced to the standard infinite horizon expected discounted reward criteria or expected total reward criteria [6, 11, 12, 14, 24].

(b) Note that the case without discount, that is, $\alpha(i,a) \equiv 0$, is allowed in the context of this chapter; see Remark 11.2.5 for further details.

(c) When the constraint constant γ in (11.1) is sufficiently large so that $U = \Pi$, then the constrained first passage optimization problem (recall Definition 11.2.3) is reduced to the usual unconstrained first passage optimization problems [13, 15, 19, 21, 22, 28].

(d) In real situations, the target set B usually represents the set of failure states of a system, and thus τ_B denotes the working life (functioning life) of the system. Therefore, our aim is to maximize the expected rewards $V_r(\mu,\pi)$ obtained before the system fails, subject to the associated costs $V_c(\mu,\pi)$ incurred before the failure of the system is not more than some constraint constant γ. In particular, if the reward function rate $r(i,a) \equiv 1$ and the discount factor $\alpha(i,a) \equiv 0$, our aim is then reduced to maximizing the expected working life of the system, subject to the associated costs $V_c(\mu,\pi)$ incurred before the failure of the system are not more than some constraint constant γ.

To obtain the existence of a constrained optimal policy, we need several sets of conditions.

Assumption 11.2.3 *There exist constants $M > 0$, $0 < \beta < 1$, and a weight function $w \ge 1$ on E such that for every $i \in B^c := E - B$,*

(a) $\sup_{a\in A(i)} |\widetilde{r}(i,a)| \le Mw(i)$, *and* $\sup_{a\in A(i)} |\widetilde{c}(i,a)| \le Mw(i)$ *for all $a \in A(i)$, where*

$$\widetilde{r}(i,a) := r(i,a)\int_0^\infty \mathrm{e}^{-\alpha(i,a)t}(1-D(t \mid i,a))\mathrm{d}t,$$

$$\widetilde{c}(i,a) := c(i,a)\int_0^\infty \mathrm{e}^{-\alpha(i,a)t}(1-D(t \mid i,a))\mathrm{d}t, \ \textit{and}$$

$$D(t \mid i,a) := Q(t,E \mid i,a).$$

(b) $\sup_{a\in A(i)}\sum_{j\in B^c} w(j)m(j \mid i,a) \le \beta w(i)$, *where* $m(j \mid i,a) := \int_0^\infty \mathrm{e}^{-\alpha(i,a)t}Q(\mathrm{d}t, j \mid i,a)$.

Remark 11.2.5. (a) In fact, Assumption 11.2.3 is a condition that ensures the first passage criteria (11.5) and (11.6) to be finite and the dynamic programming operators to be contracting; see Lemmas 11.3.1–11.3.2 below.

(b) Assumption 11.2.3(a) shows that the cost function is allowed to have neither upper nor lower bounds, while the ones in the existing works [3–5, 7, 8, 12] for the standard constrained expected discount criteria are assumed to be bounded or nonnegative (bounded below).

(c) Note that the case without discount, that is, "$\alpha(i,a) \equiv 0$", is allowed in Assumption 11.2.3. In this case, Assumption 11.2.3(b) is reduced to that there exists a constant $0 < \beta < 1$ such that

$$\sup_{a\in A(i)}\sum_{j\in B^c} w(j)P(j \mid i,a) \le \beta w(i)\ \forall\, i \in B^c \ (\text{with } P(j \mid i,a) := Q(\infty, j \mid i,a)), \tag{11.7}$$

which can be still verified. This fact is due to that the restrictions in Assumption 11.2.3(b) are imposed on the data of the set B^c rather than the entire space E. However, if the restrictions in Assumption 11.2.3(b) are imposed on the data of the entire space E, that is, there exists a constant $0 < \beta < 1$ such that

$$\sup_{a\in A(i)}\sum_{j\in E} w(j)P(j \mid i,a) \le \beta w(i)\ \ \forall\, i \in E, \tag{11.8}$$

then (11.8) fails to hold itself. Indeed, by taking $\inf_i w(i)$ in the two sides of (11.8), we can conclude from (11.8) that "$\beta \ge 1$", which leads to a contradiction with "$0 < \beta < 1$".

Assumption 11.2.4 *(a) For each $i \in B^c$, $A(i)$ is compact.*

(b) The functions $\widetilde{r}(i,a)$, $\widetilde{c}(i,a)$, and $m(j \mid i,a)$ defined in Assumption 11.2.3 are continuous in $a \in A(i)$ for each fixed $i, j \in B^c$, respectively.

(c) The function $\sum_{j\in B^c} w(j)m(j \mid i,a)$ is continuous in $a \in A(i)$, with w as in Assumption 11.2.3.

Remark 11.2.6. Assumption 11.2.4 is the compactness-continuity conditions for the first passage criteria, which is similar to the standard compactness-continuity conditions for discount and average criteria; see, for instance, Beutler and Ross [3], Guo and Hernández-Lerma [7, 8]. The difference between them lies in that the former only imposes restrictions on the data of the set B^c, while the latter focus on the data of the entire space E.

Assumption 11.2.5 *(a)* $\sum_{j\in B^c} w(j)\mu(j) < \infty$.
(b) $U_0 := \{\pi \in \Pi \mid V_c(\mu,\pi) < \gamma\} \neq \emptyset$.

Remark 11.2.7. (a) Assumption 11.2.5(a) is a condition on the "tails" of the initial distribution μ, whereas Assumption 11.2.5(b) is a Slater-like hypothesis, typical for constrained problems; see, for instance, Beutler and Ross [3], Guo and Hernández-Lerma [7, 8], and Zhang and Guo [29].

(b) It should be noted that the conditions in Assumptions 11.2.3–11.2.5 are all imposed on the data of the set B^c rather than the entire space E and thus can be fulfilled in greater generality.

Our main result is stated as following.

Theorem 11.2.1. *Suppose that Assumptions 11.2.1–11.2.5 hold. Then there exists a constrained optimal policy that may be a stationary policy or a randomized stationary policy which differ in at most one state; that is, there exist two stationary policies f^1, f^2, a state $i^* \in B^c$, and a number $p \in [0,1]$ such that $f^1(i) = f^2(i)$ for all $i \neq i^*$ and, in addition, the randomized stationary policy $\varphi^p(\cdot \mid i)$ is constrained optimal, where*

$$\varphi^p(a \mid i) = \begin{cases} p, & \text{if } a = f^1(i^*), \\ 1-p, & \text{if } a = f^2(i^*), \\ 1, & \text{if } a = f^1(i) = f^2(i), i \neq i^*. \end{cases} \tag{11.9}$$

Proof. See Sect. 11.4. □

11.3 Technical Preliminaries

This section provides some technical preliminaries necessary for the proof of Theorem 11.2.1 in Sect. 11.4.

To begin with, we define the w-norm for every real-valued function u on E by

$$\|u\|_w := \sup_{i\in E} |u(i)|/w(i),$$

where w is the so-called weight function on E as in Assumption 11.2.3. Let

$$\mathbb{B}_w(E) := \{u : \|u\|_w < \infty\}$$

be the space of w-bounded functions on E.

Lemma 11.3.1. *Suppose that Assumptions 11.2.1 and 11.2.3 hold. Then:*

(a) For each $i \in E$ and $\pi \in \Pi$,

$$|V_r(i,\pi)| \le Mw(i)/(1-\beta), \quad |V_c(i,\pi)| \le Mw(i)/(1-\beta).$$

Hence, $V_r(\cdot,\pi) \in \mathbb{B}_w(E)$, and $V_c(\cdot,\pi) \in \mathbb{B}_w(E)$.

(b) For all $i \in E$, $\pi \in \Pi$, and $u \in \mathbb{B}_w(E)$,

$$\lim_{n\to\infty} E_i^\pi \left[\prod_{k=0}^{n-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}} \mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}} u(J_n) \right] = 0,$$

where $\mathbb{1}_D$ is the indicator function on a set D.

Proof. (a) By the definition of $V_r(i,\pi)$, we see that $V_r(i,\pi)$ can be expressed as below:

$$\begin{aligned}
&V_r(i,\pi)\\
&\quad = E_i^\pi \left[\int_0^{\tau_B} \mathrm{e}^{-\int_0^t \alpha(Z(u),W(u))\mathrm{d}u} r(Z(t),W(t))\mathrm{d}t \right]\\
&\quad = E_i^\pi \left[\int_0^{\infty} \mathrm{e}^{-\int_0^t \alpha(Z(u),W(u))\mathrm{d}u} \mathbb{1}_{\{\tau_B>t\}} r(Z(t),W(t))\mathrm{d}t \right]\\
&\quad = E_i^\pi \left[\sum_{n=0}^{\infty} \int_{T_n}^{T_{n+1}} \mathrm{e}^{-\int_0^t \alpha(Z(u),W(u))\mathrm{d}u} \mathrm{d}t\, \mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}} r(J_n,A_n) \right]\\
&\quad = E_i^\pi \left[\sum_{n=0}^{\infty} \int_{0}^{X_{n+1}} \mathrm{e}^{-\int_0^{T_n+t} \alpha(Z(u),W(u))\mathrm{d}u} \mathrm{d}t\, \mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}} r(J_n,A_n) \right]\\
&\quad = E_i^\pi \left[\sum_{n=0}^{\infty} \mathrm{e}^{-\int_0^{T_n} \alpha(Z(u),W(u))\mathrm{d}u} \mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}} r(J_n,A_n) \right.\\
&\qquad \left. \int_0^{X_{n+1}} \mathrm{e}^{-\int_{T_n}^{T_n+t} \alpha(Z(u),W(u))\mathrm{d}u} \mathrm{d}t \right]\\
&\quad = E_i^\pi \left[\sum_{n=0}^{\infty} \prod_{k=0}^{n-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}} \mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}} r(J_n,A_n) \int_0^{X_{n+1}} \mathrm{e}^{-\alpha(J_n,A_n)t} \mathrm{d}t \right]\\
&\quad = E_i^\pi \left[\sum_{n=0}^{\infty} E_i^\pi \left[\prod_{k=0}^{n-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}} \mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}} r(J_n,A_n) \right.\right.\\
&\qquad \left.\left. \times \int_0^{X_{n+1}} \mathrm{e}^{-\alpha(J_n,A_n)t} \mathrm{d}t \,|X_0,J_0,A_0,\ldots,X_n,J_n,A_n \right] \right]\\
&\quad = E_i^\pi \left[\sum_{n=0}^{\infty} \prod_{k=0}^{n-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}} \mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}} r(J_n,A_n) \right.\\
&\qquad \left. \times E_i^\pi \left[\int_0^{X_{n+1}} \mathrm{e}^{-\alpha(J_n,A_n)t} \mathrm{d}t \,|X_0,J_0,A_0,\ldots,X_n,J_n,A_n \right] \right]\\
&\quad = E_i^\pi \left[\sum_{n=0}^{\infty} \prod_{k=0}^{n-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}} \mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}} r(J_n,A_n) \right.
\end{aligned}$$

$$\times \int_0^\infty \mathrm{e}^{-\alpha(J_n,A_n)t}(1-D(t \mid J_n,A_n))\mathrm{d}t\Bigg]$$

$$= E_i^\pi\left[\sum_{n=0}^{\infty}\prod_{k=0}^{n-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}\widetilde{r}(J_n,A_n)\right], \tag{11.10}$$

where the third equality follows from Assumption 11.2.1 and the ninth equality is due to the property (11.4).

We now show that for each $n = 0, 1, \ldots,$

$$E_i^\pi\left[\prod_{k=0}^{n-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}w(J_n)\right] \le \beta^n w(i). \tag{11.11}$$

Indeed, (11.11) is trivial for $n = 0$. Now, for $n \ge 1$, it follows from the property (11.4) and Assumption 11.2.3(b) that

$$\begin{aligned}
&E_i^\pi\left[\prod_{k=0}^{n-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}w(J_n)\right]\\
&= E_i^\pi\Bigg[E_i^\pi\Bigg[\prod_{k=0}^{n-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}w(J_n)\\
&\qquad \mid T_0,J_0,A_0,\ldots,T_{n-1},J_{n-1},A_{n-1}\Bigg]\Bigg]\\
&= E_i^\pi\Bigg[\prod_{k=0}^{n-2}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_{n-1}\in B^c\}}E_i^\pi\Bigg[\mathrm{e}^{-\alpha(J_{n-1},A_{n-1})X_n}\mathbb{1}_{\{J_n\in B^c\}}w(J_n)\\
&\qquad \mid T_0,J_0,A_0,\ldots,T_{n-1},J_{n-1},A_{n-1}\Bigg]\Bigg]\\
&= E_i^\pi\Bigg[\prod_{k=0}^{n-2}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_{n-1}\in B^c\}}\\
&\qquad \times\sum_{j\in B^c}\int_0^\infty \mathrm{e}^{-\alpha(J_{n-1},A_{n-1})t}w(j)Q(\mathrm{d}t,j \mid J_{n-1},A_{n-1})\Bigg]\\
&= E_i^\pi\left[\prod_{k=0}^{n-2}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_{n-1}\in B^c\}}\sum_{j\in B^c}w(j)m(j \mid J_{n-1},A_{n-1})\right]\\
&\le \beta E_i^\pi\left[\prod_{k=0}^{n-2}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_{n-1}\in B^c\}}w(J_{n-1})\right].
\end{aligned} \tag{11.12}$$

Iterating (11.12) yields (11.11).

Moreover, observe that Assumption 11.2.3(a) and (11.11) gives

$$\begin{aligned} &E_i^\pi\left[\prod_{k=0}^{n-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}|\widetilde{r}(J_n,A_n)|\right]\\ &\le ME_i^\pi\left[\prod_{k=0}^{n-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}w(J_n)\right]\\ &\le M\beta^n w(i), \end{aligned}$$

which together with (11.10) yields

$$\begin{aligned} |V_r(i,\pi)| &\le \sum_{n=0}^{\infty}E_i^\pi\left[\prod_{k=0}^{n-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}|\widetilde{r}(J_n,A_n)|\right]\\ &\le \sum_{n=0}^{\infty}M\beta^n w(i) = Mw(i)/(1-\beta). \end{aligned}$$

Thus, we get

$$\sup_{i\in E}|V_r(i,\pi)|/w(i) \le Mw(i)/(1-\beta),$$

which shows that $V_r(\cdot,\pi)\in\mathbb{B}_w(E)$.

Similarly, the conclusion for $V_c(\cdot,\pi)$ can be obtained.

(b) Since $|u(i)|\le\|u\|_w w(i)$ for all $i\in E$, it follows from (11.11) above that

$$\begin{aligned} &E_i^\pi\left[\prod_{k=0}^{n-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}|u(J_n)|\right]\\ &\le\|u\|_w E_i^\pi\left[\prod_{k=0}^{n-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}w(J_n)\right]\le\|u\|_w\beta^n w(i), \end{aligned}$$

and so part (b) follows. □

Remark 11.3.8. In fact, Lemma 11.3.1 here for first passage criteria in SMDPs is similar to Lemma 3.1 in Huang and Guo [15]. The main difference between them is due to that the discount factor $\alpha(i,a)$ here is state-action dependent, and the reward rate here is possibly unbounded (while the ones in Huang and Guo [15] are not).

For $\varphi\in\Phi$, we define the dynamic programming operators H^φ and H on $\mathbb{B}_w(E)$ as follows: for $u\in\mathbb{B}_w(E)$, if $i\in B^c$,

$$H^{\varphi}u(i) := \int_{a\in A(i)} \left[\widetilde{r}(i,a) + \sum_{j\in B^c} u(j)m(j \mid i,a)\right]\varphi(\mathrm{d}a \mid i), \tag{11.13}$$

$$Hu(i) := \sup_{a\in A(i)} \left[\widetilde{r}(i,a) + \sum_{j\in B^c} u(j)m(j \mid i,a)\right], \tag{11.14}$$

and if $i \in B$, $H^{\varphi}u(i) = Hu(i) := 0$.

Lemma 11.3.2. *Suppose that Assumptions 11.2.1 and 11.2.3 hold. Then:*

(a) For each $\varphi \in \Phi$, $V_r(\cdot,\varphi)$ is the unique solution in $\mathbb{B}_w(E)$ to the equation

$$V_r(i,\varphi) = H^{\varphi}V_r(i,\varphi)\ \forall i \in E.$$

(b) If, in addition, Assumption 11.2.4 also holds, $V_r^*(i) := \sup_{\pi\in\Pi} V_r(i,\pi)$ is the unique solution in $\mathbb{B}_w(E)$ to equation

$$V_r^*(i) = HV_r^*(i)\ \forall i \in E.$$

Moreover, there exists an $f^* \in \mathbb{F}$ such that $V_r^*(i) = H^{f^*}V_r^*(i)$, and such a policy $f^* \in \mathbb{F}$ satisfies $V_r(i,f^*) = V_r^*(i)$ for every $i \in E$.

Proof. (a) From Lemma 11.3.1, it is clear that $V_r(\cdot,\varphi) \in \mathbb{B}_w(E)$. We now establish the equation $V_r(i,\varphi) = H^{\varphi}V_r(i,\varphi)$. It is obviously true when $i \in B$, and for $i \in B^c$, by (11.10), we have

$$\begin{aligned}
&V_r(i,\varphi)\\
&= E_i^{\varphi}\left[\sum_{n=0}^{\infty}\prod_{k=0}^{n-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}\widetilde{r}(J_n,A_n)\right]\\
&= E_i^{\varphi}\left[E_i^{\varphi}\left[\sum_{n=0}^{\infty}\prod_{k=0}^{n-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}\widetilde{r}(J_n,A_n) \mid T_0,J_0,A_0,T_1,J_1\right]\right]\\
&= E_i^{\varphi}\left[\mathbb{1}_{\{J_0\in B^c\}}\widetilde{r}(J_0,A_0) + \mathrm{e}^{-\alpha(J_0,A_0)X_1}\mathbb{1}_{\{J_0\in B^c,J_1\in B^c\}}E_i^{\varphi}\left[\sum_{n=1}^{\infty}\prod_{k=1}^{n-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\right.\right.\\
&\qquad \left.\left.\mathbb{1}_{\{J_2\in B^c,\ldots,J_n\in B^c\}}\widetilde{r}(J_n,A_n) \mid T_0,J_0,A_0,T_1,J_1\right]\right]\\
&= \int_{a\in A(i)}\varphi(\mathrm{d}a \mid i)\left[\widetilde{r}(i,a) + \sum_{j\in B^c}\int_0^{\infty}\mathrm{e}^{-\alpha(i,a)t}Q(\mathrm{d}t,j \mid i,a)E_i^{\varphi}\left[\sum_{n=1}^{\infty}\prod_{k=1}^{n-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\right.\right.\\
&\qquad \left.\left.\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}\widetilde{r}(J_n,A_n) \mid T_0=0,J_0=i,A_0=a,T_0=t,J_1=j\right]\right]
\end{aligned}$$

$$= \int_{a\in A(i)} \varphi(\mathrm{d}a \mid i)\left[\widetilde{r}(i,a) + \sum_{j\in B^c} m(j \mid i,a) E_j^{\varphi}\left[\sum_{n=0}^{\infty}\prod_{k=0}^{n-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}} \right.\right.$$

$$\left.\left. \mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}\widetilde{r}(J_n,A_n)\right]\right]$$

$$= \int_{a\in A(i)} \varphi(\mathrm{d}a \mid i)\left[\widetilde{r}(i,a) + \sum_{j\in B^c} m(j \mid i,a) V_r(j,\varphi)\right],$$

where the fifth equality is due to the properties (11.2)–(11.4) and that policy φ is Markov. Hence, we obtain that $V_r(i,\varphi) = H^{\varphi}V_r(i,\varphi),\ i \in E$.

To complete the proof, we need only show that H^{φ} is a contraction from $\mathbb{B}_w(E)$ to $\mathbb{B}_w(E)$, and thus H^{φ} has a unique fixed point in $\mathbb{B}_w(E)$. Indeed, for an arbitrary function $u \in \mathbb{B}_w(E)$, by Assumption 11.2.3 it is clear that

$$\begin{aligned}
&|H^{\varphi}u(i)| \\
&= \left|\int_{a\in A(i)} \left[\widetilde{r}(i,a) + \sum_{j\in B^c} u(j)m(j \mid i,a)\right]\varphi(\mathrm{d}a \mid i)\right| \\
&\le \int_{a\in A(i)} \left[|\widetilde{r}(i,a)| + \sum_{j\in B^c} |u(j)|m(j \mid i,a)\right]\varphi(\mathrm{d}a \mid i) \\
&\le \int_{a\in A(i)} \left[Mw(i) + \|u\|_w \beta w(i)\right]\varphi(\mathrm{d}a \mid i) \\
&= (M + \beta\|u\|_w)w(i)\ \forall i \in B^c,
\end{aligned}$$

which implies that $H^{\varphi}u \in \mathbb{B}_w(E)$, that is, H^{φ} maps $\mathbb{B}_w(E)$ to itself. Moreover, for any $u, u' \in \mathbb{B}_w(E)$, we have

$$\begin{aligned}
&|H^{\varphi}u(i) - H^{\varphi}u'(i)| \\
&= \left|\int_{a\in A(i)} \left[\sum_{j\in B^c} (u(j) - u'(j))m(j \mid i,a)\right]\varphi(\mathrm{d}a \mid i)\right| \\
&\le \int_{a\in A(i)} \left[\sum_{j\in B^c} |u(j) - u'(j)|m(j \mid i,a)\right]\varphi(\mathrm{d}a \mid i) \\
&\le \int_{a\in A(i)} \left[\|u - u'\|_w \beta w(i)\right]\varphi(\mathrm{d}a \mid i) \\
&= \beta\|u - u'\|_w w(i)\ \forall i \in B^c.
\end{aligned}$$

Hence, $\|H^{\varphi}u - H^{\varphi}u'\|_w \le \beta\|u - u'\|_w$, and thus H^{φ} is a contraction from $\mathbb{B}_w(E)$ to itself. By Banach's Fixed Point Theorem, H^{φ} has a unique fixed point in $\mathbb{B}_w(E)$, and so the proof is achieved.

(b) Under Assumption 11.2.4, using a similar manner to the proof of part (a) yields that H is a contraction from $\mathbb{B}_w(E)$ to itself, and thus, by Banach's Fixed Point Theorem, H has a unique fixed point u^* in $\mathbb{B}_w(E)$, that is, $Hu^* = u^*$. Hence, to prove part (b), we need to show that: (b_1) $V_r^* \in \mathbb{B}_w(E)$, with w-norm $\|V_r^*\| \leq M/(1-\beta)$. (b_2) $V_r^* = u^*$.

In fact, (b_1) is an immediate result of Lemma 11.3.1(a). Thus, it remains to prove (b_2). To this end, we show that $u^* \leq V_r^*$ and $u^* \geq V_r^*$ as below, respectively. It is clear that $u^*(i) = V_r^*(i) = 0$ for every $i \in B$. Hence, in the following, we restrict our argument to the case of $i \in B^c$.

(i) This part is to show that $u^* \leq V_r^*$. By the equality $u^* = Hu^*$ and the measurable selection theorem [10, Proposition D.5, p.182], there exists an $f \in \mathbb{F}$ such that

$$u^*(i) = \widetilde{r}(i,f) + \sum_{j \in B^c} u^*(j) m(j \mid i,f) \ \forall i \in B^c. \tag{11.15}$$

Iteration of (11.15) yields

$$\begin{aligned} u^*(i) = E_i^f \left[\sum_{m=0}^{n-1} \prod_{k=0}^{m-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}} \mathbb{1}_{\{J_0 \in B^c, \ldots, J_m \in B^c\}} \widetilde{r}(J_m,f) \right] \\ + E_i^f \left[\prod_{k=0}^{n-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}} \mathbb{1}_{\{J_0 \in B^c, \ldots, J_n \in B^c\}} u^*(J_n) \right] \ \forall i \in B^c, \ n = 1,2\ldots, \end{aligned}$$

and letting $n \to \infty$ we get, by Lemma 11.3.1(b),

$$u^*(i) = E_i^f \left[\sum_{m=0}^{\infty} \prod_{k=0}^{m-1} \mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}} \mathbb{1}_{\{J_0 \in B^c, \ldots, J_m \in B^c\}} \widetilde{r}(J_m,f) \right] = V_r(i,f) \ \forall i \in B^c.$$

Thus, by the definition of V_r^*, we see that $u^* \leq V_r^*$.

(ii) This part is to show that $u^* \geq V_r^*$. Note that $u^* = Hu^*$ implies that

$$u^*(i) \geq \widetilde{r}(i,a) + \sum_{j \in B^c} u^*(j) m(j \mid i,a) \ \forall i \in B^c, \ a \in A(i), \tag{11.16}$$

which gives

$$\mathbb{1}_{\{J_n \in B^c\}} u^*(J_n) \geq \mathbb{1}_{\{J_n \in B^c\}} \widetilde{r}(J_n,A_n) + \mathbb{1}_{\{J_n \in B^c\}} \sum_{j \in B^c} u^*(j) m(j \mid J_n,A_n) \ \forall n \geq 0. \tag{11.17}$$

Hence, for any initial state $i \in B^c$ and policy $\pi \in \Pi$, using properties (11.2)–(11.4) yields

$$\mathbb{1}_{\{J_n\in B^c\}}u^*(J_n) \geq E_i^\pi\Big[\mathbb{1}_{\{J_n\in B^c\}}\widetilde{r}(J_n,A_n)+\mathrm{e}^{-\alpha(J_n,A_n)X_{n+1}}\mathbb{1}_{\{J_n\in B^c,J_{n+1}\in B^c\}}$$
$$\times u^*(J_{n+1})\mid T_0,J_0,A_0,\ldots,T_n,J_n,A_n\Big]\forall n=0,1\ldots,$$

which gives

$$\prod_{k=0}^{n-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}u^*(J_n)$$
$$\geq E_i^\pi\Big[\prod_{k=0}^{n-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}\widetilde{r}(J_n,A_n)$$
$$+\prod_{k=0}^{n}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_{n+1}\in B^c\}}u^*(J_{n+1})\mid T_0,J_0,A_0,\ldots,T_n,J_n,A_n\Big]$$
$$\forall n=0,1\ldots.$$

Therefore, taking expectation E_i^π and summing over $m=0,1\ldots,n-1$, we obtain

$$u^*(i)\geq E_i^\pi\Big[\sum_{m=0}^{n-1}\prod_{k=0}^{m-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_m\in B^c\}}\widetilde{r}(J_m,A_m)\Big]$$
$$+E_i^\pi\Big[\prod_{k=0}^{n-1}\mathrm{e}^{-\alpha(J_k,A_k)X_{k+1}}\mathbb{1}_{\{J_0\in B^c,\ldots,J_n\in B^c\}}u^*(J_n)\Big],\quad\forall n=1,2\ldots.$$

Finally, letting $n\to\infty$ in the latter inequality and using Lemma 11.3.1(b), it follows that

$$u^*(i)\geq V_r(i,\pi)$$

so that, as i and π were arbitrary, we conclude that $u^*\geq V_r^*$.

Combining (i) with (ii) yields that $u^*=V_r^*$, and thus we have $V_r^*=HV_r^*$.

Finally, it follows from $V_r^*=HV_r^*$ and the measurable selection theorem that there exists an $f^*\in\mathbb{F}$ such that $V_r^*=H^{f^*}V_r^*$. This fact together with part (a) implies that $V_r^{f^*}=V_r^*$. □

Remark 11.3.9. Note that Lemma 11.3.2 also holds for the case of the expected first passage cost V_c accordingly.

Note that $\mathbb{F}$ can be written as the product space $\mathbb{F}=\prod_{i\in E}A(i)$. Hence, by Assumption 11.2.4(a) and Tychonoff's theorem, $\mathbb{F}$ is a compact metric space.

Lemma 11.3.3. *Suppose that Assumptions 11.2.1–11.2.4 and 11.2.5(a) hold. Then the functions $V_r(\mu,f)$ and $V_c(\mu,f)$ are continuous in $f\in\mathbb{F}$.*

Proof. We only prove the continuity of $V_r(\mu,f)$ in $f\in\mathbb{F}$ because the other case is similar. Let $f_n\to f$ as $n\to\infty$ and fix any $i\in E$. Let $v(i):=\limsup_{n\to\infty}V_r(i,f_n)$.

Then, by Theorem 4.4 in [1], there exists a subsequence $\{V_r(i,f_{n_m})\}$ (depending on i) of $\{V_r(i,f_n)\}$ such that $V_r(i,f_{n_m}) \to v(i)$ as $m \to \infty$. Then, by Lemma 11.3.1(a), we have $V_r(j,f_{n_m}) \in [-Mw(j)/(1-\beta), Mw(j)/(1-\beta)]$ for all $j \in E$ and $m \geq 1$, and so $V_r(\cdot,f_{n_m})$ is in the product space $\prod_{j\in E}[-Mw(j)/(1-\beta), Mw(j)/(1-\beta)]$ for each $m \geq 1$. Since E is denumerable, the Tychonoff theorem shows that the space $\prod_{j\in E}[-Mw(j)/(1-\beta), Mw(j)/(1-\beta)]$ is compact, and thus there exists a subsequence $\{V_r(\cdot,f_{n_k})\}$ of $\{V_r(\cdot,f_{n_m})\}$ converging to some point u in $\prod_{j\in E}[-Mw(j)/(1-\beta), Mw(j)/(1-\beta)]$, that is, $\lim_{k\to\infty} V_r(j,f_{n_k}) = u(j)$ for all $j \in E$, which, together with $f_n \to f$ and $\lim_{m\to\infty} V_r(i,f_{n_m}) = v(i)$, implies that

$$v(i) = u(i), \ \lim_{k\to\infty} V_r(j,f_{n_k}) = u(j), \ \text{and} \lim_{k\to\infty} f_{n_k}(j) = f(j), \ \text{ for all } j \in E. \quad (11.18)$$

Moreover, by Lemma 11.3.1(a), we have

$$|u(j)| \leq Mw(j)/(1-\beta), \quad \text{for all } j \in E, \quad (11.19)$$

which implies that $u \in \mathbb{B}_w(E)$.

On the other hand, for $k \geq 1$ and the given $i \in B^c$, by Lemma 11.3.2(a), we have

$$V_r(i,f_{n_k}) = \widetilde{r}(i,f_{n_k}) + \sum_{j\in B^c} V_r(j,f_{n_k}) m(j \mid i,f_{n_k}). \quad (11.20)$$

Then, under Assumptions 11.2.3 and 11.2.4, from (11.18)–(11.20) and Lemma 8.3.7 (i.e., the Generalized Dominated Convergence Theorem) in [11], we get

$$u(i) = \widetilde{r}(i,f) + \sum_{j\in B^c} u(j) m(j \mid i,f). \quad (11.21)$$

Thus, by Lemma 11.3.2(a) and (11.18), we conclude that

$$\limsup_{n\to\infty} V_r(i,f_n) = v(i) = u(i) = V_r(i,f). \quad (11.22)$$

Similarly, we can prove that

$$\liminf_{n\to\infty} V_r(i,f_n) = V_r(i,f),$$

which together with (11.22) implies that

$$\limsup_{n\to\infty} V_r(i,f_n) = \liminf_{n\to\infty} V_r(i,f_n) = V_r(i,f),$$

and so

$$\lim_{n\to\infty} V_r(i,f_n) = V_r(i,f). \quad (11.23)$$

Moreover, noting that $V_r(i,f_n) = V_r(i,f) = 0$ for every $i \in B$ and $n \geq 0$, it follows from Assumption 11.2.5(a) and Lemma 8.3.7 in [11] again that

$$\begin{aligned}\lim_{n\to\infty} V_r(\mu,f_n) &= \lim_{n\to\infty}\sum_{i\in E}[V_r(i,f_n)]\mu(i) = \lim_{n\to\infty}\sum_{i\in B^c}[V_r(i,f_n)]\mu(i)\\ &= \sum_{i\in B^c}[\lim_{n\to\infty} V_r(i,f_n)]\mu(i) = \sum_{i\in B^c} V_r(i,f)\mu(i) = V_r(\mu,f), \qquad (11.24)\end{aligned}$$

which gives the stated result: $V_r(\mu,f_n) \to V_r(\mu,f)$, as $n \to \infty$. □

To analyze the constrained control problem (recall Definition 11.2.3), we introduce a Lagrange multiplier $\lambda \geq 0$ as follows. For each $i \in E$ and $a \in A(i)$, let

$$b^\lambda(i,a) := r(i,a) - \lambda c(i,a). \qquad (11.25)$$

Furthermore, for each policy $\pi \in \Pi$ and $i \in E$, let

$$V_{b^\lambda}(i,\pi) := E_i^\pi\left[\int_0^{\tau_B} \mathrm{e}^{-\int_0^t \alpha(Z(u),W(u))\mathrm{d}u} b^\lambda(Z(t),W(t))\mathrm{d}t\right], \qquad (11.26)$$

$$V_{b^\lambda}(\mu,\pi) := \sum_{j\in E} V_{b^\lambda}(j,\pi)\mu(j), \qquad (11.27)$$

$$V^*_{b^\lambda}(i) := \sup_{\pi\in\Pi} V_{b^\lambda}(i,\pi), V^*_{b^\lambda}(\mu) := \sup_{\pi\in\Pi} V_{b^\lambda}(\mu,\pi). \qquad (11.28)$$

Remark 11.3.10. Notice that, for each $i \in B$, $V_{b^\lambda}(i,\pi) = 0$. Therefore, we have

$$V_{b^\lambda}(\mu,\pi) = \sum_{j\in B^c} V_{b^\lambda}(j,\pi)\mu(j), \qquad V^*_{b^\lambda}(\mu) = \sum_{j\in B^c} V^*_{b^\lambda}(j)\mu(j).$$

Under Assumptions 11.2.1–11.2.4, by Lemma 11.3.2(b), we have

$$V^*_{b^\lambda}(i) = \begin{cases} 0, & \text{for } i \in B,\\ \sup\limits_{a\in A(i)}\left[\widetilde{b^\lambda}(i,a) + \sum\limits_{j\in B^c} V^*_{b^\lambda}(j)m(j \mid i,a)\right], & \text{for } i \in B^c, \end{cases} \qquad (11.29)$$

where $\widetilde{b^\lambda}(i,a) := b^\lambda(i,a)\int_0^\infty \mathrm{e}^{-\alpha t}(1 - D(t \mid i,a))\mathrm{d}t$. Moreover, for each $i \in E$, the maximum in (11.29) is realized by some $a \in A(i)$, that is,

$$A^*_\lambda(i) := \begin{cases} A(i), & \text{for } i \in B,\\ \left\{a \in A(i) \mid V^*_{b^\lambda}(i) = \widetilde{b^\lambda}(i,a) + \sum_{j\in B^c} V^*_{b^\lambda}(j)m(j \mid i,a)\right\}, & \text{for } i \in B^c \end{cases} \qquad (11.30)$$

is *nonempty*. In other words, the following sets

$$\mathbb{F}^*_\lambda := \left\{ f \in \mathbb{F} \mid f(i) \in A^*_\lambda(i)\ \forall i \in E \right\} \tag{11.31}$$

and

$$\Phi^\lambda := \left\{ \varphi \in \Phi \mid \varphi(A^*_\lambda(i) \mid i) = 1\ \forall i \in E \right\} \tag{11.32}$$

are *nonempty*.

Next lemma reveals that Φ^λ is convex.

Lemma 11.3.4. *Under Assumptions 11.2.1–11.2.4, the set Φ^λ is convex.*

Proof. For each $\varphi_1, \varphi_2 \in \Phi^\lambda$, and $p \in [0,1]$, let

$$\varphi^p(\cdot \mid i) := p\varphi_1(\cdot \mid i) + (1-p)\varphi_2(\cdot \mid i),\ \forall i \in E. \tag{11.33}$$

Hence, $\varphi^p(A^*_\lambda(i) \mid i) = p\varphi_1(A^*_\lambda(i) \mid i) + (1-p)\varphi_2(A^*_\lambda(i) \mid i) = 1$, and so Φ^λ is convex. □

Notation. For each $\lambda \geq 0$, we take an arbitrary, but fixed policy $f^\lambda \in \mathbb{F}^*_\lambda$, and denote $V_r(\mu, f^\lambda)$, $V_c(\mu, f^\lambda)$, and $V_{b^\lambda}(\mu, f^\lambda)$ by $V_r(\lambda)$, $V_c(\lambda)$, and $V_b(\lambda)$, respectively. By Lemma 11.3.2, we have that $V_{b^\lambda}(i, f) = V^*_{b^\lambda}(i)$ for all $i \in E$ and $f \in \mathbb{F}^*_\lambda$. Hence, $V_b(\lambda) = V_{b^\lambda}(\mu, f^\lambda) = V^*_{b^\lambda}(\mu)$.

Lemma 11.3.5. *If Assumptions 11.2.3–11.2.4 and 11.2.5(a) hold, then $V_c(\lambda)$ is nonincreasing in $\lambda \in [0, \infty)$.*

Proof. By (11.5), (11.6), and (11.25)–(11.26) for each $\pi \in \Pi$, we obtain

$$V_{b^\lambda}(\mu, \pi) = V_r(\mu, \pi) - \lambda V_c(\mu, \pi)\ \forall \lambda \geq 0.$$

Moreover, noting that $V_b(\lambda) = V_{b^\lambda}(\mu, f^\lambda) = V^*_{b^\lambda}(\mu)$ for all $\lambda \geq 0$ and $f^\lambda \in \mathbb{F}^*_\lambda$, we have, for any $h > 0$,

$$\begin{aligned} -hV_c(\lambda) &= V_{b^{\lambda+h}}(\mu, f^\lambda) - V_b(\lambda) \\ &\leq V_b(\lambda + h) - V_b(\lambda) \\ &\leq V_b(\lambda + h) - V_{b^\lambda}(\mu, f^{\lambda+h}) = -hV_c(\lambda + h), \end{aligned}$$

which gives that

$$-hV_c(\lambda) \leq -hV_c(\lambda + h).$$

Hence, $V_c(\lambda)$ is nonincreasing in $\lambda \in [0, \infty)$. □

Lemma 11.3.6. *Suppose that Assumptions 11.2.1–11.2.4 hold. If $\lim_{k\to\infty} \lambda_k = \lambda$, and $f^{\lambda_k} \in \mathbb{F}^*_{\lambda_k}$ is such that $\lim_{k\to\infty} f^{\lambda_k} = f$, then $f \in \mathbb{F}^*_\lambda$.*

Proof. Since $f^{\lambda_k} \in \mathbb{F}^*_{\lambda_k}$, for each $i \in B^c$ and $\pi \in \Pi$, we have

$$V^*_{b^{\lambda_k}}(i) = V_r(i, f^{\lambda_k}) - \lambda_k V_c(i, f^{\lambda_k}) \geq V_{b^{\lambda_k}}(i,\pi) = V_r(i,\pi) - \lambda_k V_c(i,\pi). \quad (11.34)$$

Letting $k \to \infty$ in (11.34) and by Lemma 11.3.3, we obtain

$$V_{b^\lambda}(i,f) \geq V_{b^\lambda}(i,\pi) \ \ \forall i \in B^c \ \text{ and } \ \pi \in \Pi,$$

which together with the fact that $A^*_\lambda(i) = A(i)$ for each $i \in B$ implies that $f \in \mathbb{F}^*_\lambda$. □

Under Assumptions 11.2.1–11.2.4 and 11.2.5(a), it follows from Lemma 11.3.5 that the following nonnegative constant

$$\overline{\lambda} := \inf\{\lambda \geq 0 \mid V_c(\lambda) \leq \gamma\} \quad (11.35)$$

is well defined.

Lemma 11.3.7. *Suppose that Assumptions 11.2.1–11.2.5 hold. Then the constant $\overline{\lambda}$ in (11.35) is finite, that is, $\overline{\lambda} \in [0,\infty)$.*

Proof. Suppose that $\overline{\lambda} = \infty$. By Assumption 11.2.5(b), there exists a policy $\pi' \in \Pi$ such that $V_c(\mu,\pi') < \gamma$. Let $d := \gamma - V_c(\mu,\pi') > 0$. Then, for any $\lambda > 0$, we have

$$V_{b^\lambda}(\mu,\pi') = V_r(\mu,\pi') - \lambda V_c(\mu,\pi') = V_r(\mu,\pi') - \lambda(\gamma - d). \quad (11.36)$$

As $\overline{\lambda} = \infty$, by (11.35) and Lemma 11.3.5, we obtain $V_c(\lambda) > \gamma$ for all $\lambda > 0$. Therefore, $V_b(\lambda) = V_r(\lambda) - \lambda V_c(\lambda) < V_r(\lambda) - \lambda\gamma$. Since $V_b(\lambda) = V^*_{b^\lambda}(\mu) \geq V_{b^\lambda}(\mu,\pi')$, from (11.36), we have

$$V_r(\lambda) - \lambda\gamma > V_b(\lambda) \geq V_{b^\lambda}(\mu,\pi') = V_r(\mu,\pi') - \lambda(\gamma - d) \ \ \forall \lambda > 0, \quad (11.37)$$

which gives

$$V_r(\lambda) > V_r(\mu,\pi') + \lambda d \quad \forall \lambda > 0. \quad (11.38)$$

On the other hand, by Lemma 11.3.1 and Assumption 11.2.5(a), we have

$$\max\Big\{|V_r(\mu,\pi')|, |V_r(\lambda)|\Big\} \leq M\Big[\sum_{j\in B^c} w(j)\mu(j)\Big]/(1-\beta) := \tilde{M} < \infty \quad (11.39)$$

for all $\lambda > 0$. The latter inequality together with (11.38) gives that

$$2\tilde{M} > \lambda d \ \ \forall \lambda > 0, \quad (11.40)$$

which is clearly a contradiction; for instance, take $\lambda = 3\tilde{M}/d > 0$. Hence, $\overline{\lambda}$ is finite. □

11.4 Proof of Theorem 11.2.1

In this section, we prove Theorem 11.2.1 by using the Lagrange approach and the following lemma.

Lemma 11.4.8. *If there exist $\lambda_0 \geq 0$ and $\pi^* \in U$ such that*

$$V_c(\mu,\pi^*) = \gamma \text{ and } V_{b^{\lambda_0}}(\mu,\pi^*) = V^*_{b^{\lambda_0}}(\mu),$$

then π^* is constrained optimal.

Proof. For any $\pi \in U$, since $V_{b^{\lambda_0}}(\mu,\pi^*) = V^*_{b^{\lambda_0}}(\mu) \geq V_{b^{\lambda_0}}(\mu,\pi)$, we have

$$V_r(\mu,\pi^*) - \lambda_0 V_c(\mu,\pi^*) \geq V_r(\mu,\pi) - \lambda_0 V_c(\mu,\pi). \tag{11.41}$$

Noting that $V_c(\mu,\pi^*) = \gamma$ and $V_c(\mu,\pi) \leq \gamma$ (by $\pi \in U$), from (11.41), we get

$$V_r(\mu,\pi^*) \geq V_r(\mu,\pi) + \lambda_0(\gamma - V_c(\mu,\pi)) \geq V_r(\mu,\pi)\ \forall \pi \in U,$$

which means that π^* is constrained optimal. □

Proof of Theorem 11.2.1. By Lemma 11.3.7, the constant $\overline{\lambda} \in [0,\infty)$. Thus, we shall consider the two cases: $\overline{\lambda} = 0$ and $\overline{\lambda} > 0$.

The case of $\overline{\lambda} = 0$: By (11.35), there exists a sequence $f^{\lambda_k} \in \mathbb{F}^*_{\lambda_k}$ such that $\lambda_k \downarrow 0$ as $k \to \infty$. Because $\mathbb{F}$ is compact, we may assume that $f^{\lambda_k} \to \tilde{f}$ without loss of generality. Thus, by Lemma 11.3.5, we have $V_c(\mu, f^{\lambda_k}) \leq \gamma$ for all $k \geq 1$, and then it follows from Lemma 11.3.3 that $\tilde{f} \in U$. Moreover, for each $\pi \in U$, we have that $V_b(\lambda_k) = V_{b^{\lambda_k}}(\mu, f^{\lambda_k}) \geq V_{b^{\lambda_k}}(\mu,\pi)$. Hence, by Lemma 11.3.1(a) and (11.39),

$$V_r(\mu, f^{\lambda_k}) - V_r(\mu,\pi) \geq \lambda_k[V_c(\mu, f^{\lambda_k}) - V_c(\mu,\pi)] \geq -2\lambda_k \tilde{M}. \tag{11.42}$$

Letting $k \to \infty$ in (11.42), by Lemma 11.3.3, we have

$$V_r(\mu,\tilde{f}) - V_r(\mu,\pi) \geq 0\ \forall \pi \in U,$$

which means that $\tilde{f}$ is a constrained optimal stationary policy.

The case of $\overline{\lambda} > 0$: First, if there is some $\lambda' \in (0,\infty)$ satisfying $V_c(\lambda') = \gamma$, then there exist an associated $f^{\lambda'} \in \mathbb{F}^*_{\lambda'}$ such that $V_c(\lambda') = V_c(\mu, f^{\lambda'}) = \gamma$, and $V^*_{b^{\lambda'}}(\mu) = V_{b^{\lambda'}}(\mu, f^{\lambda'})$. Thus, by Lemma 11.4.8, $f^{\lambda'}$ is a constrained optimal stationary policy.

Now, suppose that $V_c(\lambda) \neq \gamma$ for all $\lambda \in (0,\infty)$. Then, as $\overline{\lambda}$ is in $(0,\infty)$, there exist two nonnegative sequences $\{\lambda_k\}$ and $\{\delta_k\}$ such that $\lambda_k \uparrow \overline{\lambda}$ and $\delta_k \downarrow \overline{\lambda}$, respectively. Since $\mathbb{F}$ is compact, we may take $f^{\lambda_k} \in \mathbb{F}^*_{\lambda_k}$ and $f^{\delta_k} \in \mathbb{F}^*_{\delta_k}$ such that $f^{\lambda_k} \to f^1 \in \mathbb{F}$ and $f^{\delta_k} \to f^2 \in \mathbb{F}$, without loss of generality. By Lemma 11.3.6, we have that $f^1, f^2 \in$

$\mathbb{F}^*_{\overline{\lambda}}$. By Lemmas 11.3.3 and 11.3.4, we obtain that $V_c(\mu, f^1) \geq \gamma$ and $V_c(\mu, f^2) \leq \gamma$. If $V_c(\mu, f^1) = \gamma$ (or $V_c(\mu, f^2) = \gamma$), by Lemma 11.4.8, we have that f^1 (or f^2) is a constrained optimal stationary policy. Hence, to complete the proof, we shall consider the following case:

$$V_c(\mu, f^1) > \gamma \text{ and } V_c(\mu, f^2) < \gamma. \tag{11.43}$$

Now using f^1 and f^2, we construct a sequence of stationary policies $\{f_n\}$ as follows. For each $n \geq 1$ and $i \in E$, let

$$f_n(i) := \begin{cases} f^1(i), \text{ if } i < n; \\ f^2(i), \text{ if } i \geq n, \end{cases}$$

where, without loss of generality, the denumerable state space is assumed to be the set $\{1, 2, \ldots\}$. Obviously, $f_1 = f^2$ and $\lim_{n\to\infty} f_n = f^1$. Hence, by Lemma 11.3.3, $\lim_{n\to\infty} V_c(\mu, f_n) = V_c(\mu, f^1)$. Since $f^1, f^2 \in \mathbb{F}^*_{\overline{\lambda}}$ (just mentioned), by (11.31), we see that $f_n \in \mathbb{F}^*_{\overline{\lambda}}$ for all $n \geq 1$. As $f_1 = f^2$, by (11.43), we have $V_c(\mu, f_1) < \gamma$. If there exists n^* such that $V_c(\mu, f_{n^*}) = \gamma$, then by Lemma 11.4.8 and $f_n \in \mathbb{F}^*_{\overline{\lambda}}$, f_{n^*} a constrained optimal stationary policy. Thus, in the remainder of this section, we may assume that $V_c(\mu, f_n) \neq \gamma$ for all $n \geq 1$. If $V_c(\mu, f_n) < \gamma$ for all $n \geq 1$, $\lim_{n\to\infty} V_c(\mu, f_n) = V_c(\mu, f^1) \leq \gamma$, which is a contradiction to (11.43). Thus, there exists some $n \geq 1$ such that $V_c(\mu, f_n) > \gamma$, which together with $V_c(\mu, f_1) < \gamma$ gives the existence of some $\tilde{n}$ such that

$$V_c(\mu, f_{\tilde{n}}) < \gamma \text{ and } V_c(\mu, f_{\tilde{n}+1}) > \gamma. \tag{11.44}$$

Obviously, the stationary policies $f_{\tilde{n}}$ and $f_{\tilde{n}+1}$ differ in at most the state $\tilde{n}$. Here, it should be pointed out that $\tilde{n}$ must be in B^c. Indeed, if $\tilde{n} \in B$, we have $V_c(\tilde{n}, f_{\tilde{n}}) = V_c(\tilde{n}, f_{\tilde{n}+1}) = 0$, which implies that $V_c(\mu, f_{\tilde{n}}) = V_c(\mu, f_{\tilde{n}+1})$ and thus leads to a contradiction to (11.44).

For any $p \in [0, 1]$, using the stationary policies $f_{\tilde{n}}$ and $f_{\tilde{n}+1}$, we construct a randomized stationary policy φ^p as follows. For each $i \in E$,

$$\varphi^p(a \mid i) = \begin{cases} p, & \text{if } a = f_{\tilde{n}}(\tilde{n}) \text{ when } i = \tilde{n}, \\ 1 - p, & \text{if } a = f_{\tilde{n}+1}(\tilde{n}) \text{ when } i = \tilde{n}, \\ 1, & \text{if } a = f_{\tilde{n}}(i) \text{ when } i \neq \tilde{n}. \end{cases} \tag{11.45}$$

Since $f_{\tilde{n}}, f_{\tilde{n}+1} \in \mathbb{F}^*_{\overline{\lambda}}$, by Lemma 11.3.4, we have $V_{b^{\overline{\lambda}}}(\mu, \varphi^p) = V^*_{b^{\overline{\lambda}}}(\mu)$ for all $p \in [0, 1]$. We also have that $V_c(\mu, \varphi^p)$ is continuous in $p \in [0, 1]$. Indeed, for any $p \in [0, 1]$ and any sequence $\{p_m\}$ in $[0, 1]$ such that $\lim_{n\to\infty} p_m = p$, as in the proof of Lemma 11.3.2, we have

$$V_c(i,\varphi^{p_m}) = \sum_{a\in A(i)} \varphi^{p_m}(a \mid i)\left[\widetilde{c}(i,a) + \sum_{j\in B^c} V_c(j,\varphi^{p_m})m(j \mid i,a)\right] \ \forall i \in B^c. \quad (11.46)$$

Hence, as in the proof of Lemma 11.3.3, from (11.45) and (11.46), we obtain

$$\lim_{n\to\infty} V_c(\mu,\varphi^{p_m}) = V_c(\mu,\varphi^{p}),$$

and so $V_c(\mu,\varphi^p)$ is continuous in $p \in [0,1]$.

Finally, let $p_0 = 0$ and $p_1 = 1$. Then, $V_c(\mu,\varphi^{p_0}) = V_c(\mu,f_{\tilde{n}+1}) > \gamma$ and $V_c(\mu,\varphi^{p_1}) = V_c(\mu,f_{\tilde{n}}) < \gamma$. Therefore, by the continuity of $V_c(\mu,\varphi^p)$ in $p \in [0,1]$ there exists a $p^* \in (0,1)$ such that $V_c(\mu,\varphi^{p^*}) = \gamma$. Since $V_{b^{\bar{\lambda}}}(\mu,\varphi^{p^*}) = V^*_{b^{\bar{\lambda}}}(\mu)$, by Lemma 11.4.8, we have that φ^{p^*} is a constrained optimal stationary policy, which randomizes between the two stationary policies $f_{\tilde{n}}$ and $f_{\tilde{n}+1}$ that differ in at most the state $\tilde{n} \in B^c$. □

References

1. Aliprantis, C.D., Burkinshaw O.: Principles of real analysis. Third edition. Academic Press, Inc., San Diego, CA (1998)
2. Berument, H., Kilinc, Z., Ozlale, U.: The effects of different inflation risk premiums on interest rate spreads. Phys. A **333**, 317–324 (2004)
3. Beutler, F.J., Ross, K.W.: Time-average optimal constrained semi-Markov decision processes. Adv. in Appl. Probab. **18**, 341–359 (1986)
4. Feinberg, E.A.: Constrained semi-Markov decision processes with average rewards. Z. Oper. Res. **39**, 257–288 (1994)
5. Feinberg, E.A.: Continuous time discounted jump Markov decision processes: a discrete-event approach. Math. Oper. Res. **29**, 492–524 (2004)
6. Guo, X.P.: Constrained denumerable state non-stationary MDPs with expected total reward criterion. Acta Math. Appl. Sinica (English Ser.) **16**, 205–212 (2000)
7. Guo, X.P., Hernández-Lerma, O.: Constrained continuous-time Markov control processes with discounted criteria. Stochastic Anal. Appl. **21**, 379–399 (2003)
8. Guo, X.P., Hernández-Lerma, O.: Continuous-Time Markov Decision Processes: Theory and Applications. Springer-Verlag, Berlin Heidelberg (2009)
9. Haberman, S., Sung, J.: Optimal pension funding dynamics over infinite control horizon when stochastic rates of return are stationary. Insur. Math. Econ. **36**, 103–116 (2005)
10. Hernández-Lerma, O., Lasserre, J.B.: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer-Verlag, New York (1996)
11. Hernández-Lerma, O., Lasserre, J.B.: Further Topics on Discrete-Time Markov Control Processes. Springer-Verlag, New York (1999)
12. Hernández-Lerma, O., González-Hernández, J.: Constrained Markov control processes in Borel spaces: the discounted case. Math. Methods Oper. Res. **52**, 271–285 (2000)
13. Huang, Y.H, Guo, X.P.: Optimal risk probability for first passage models in semi-Markov decision processes. J. Math. Anal. Appl. **359**, 404–420 (2009)
14. Huang, Y.H, Guo, X.P.: Discounted semi-Markov decision processes with nonnegative costs. Acta. Math. Sinica (Chinese Series) **53**, 503–514 (2010)
15. Huang, Y.H, Guo, X.P.: First passage models for denumerable semi-Markov decision processes with nonnegative discounted costs. Acta. Math. Appl. Sinica **27**, 177–190 (2011)

16. Huang, Y.H, Guo, X.P.: Finite horizon semi-Markov decision processes with application to maintenance systems. European. J. Oper. Res. **212**, 131–140 (2011)
17. Lee, P., Rosenfield, D.B.: When to refinance a mortgage: a dynamic programming approach. European. J. Oper. Res. **166**, 266–277 (2005)
18. Limnios, N., Oprisan, J.: Semi-Markov Processes and Reliability. Birkhäuser, Boston (2001)
19. Lin, Y.L.: Continuous time first arrival target models (1)- discounted moment optimal models. Acta. Math. Appl. Sinica-Chinese Serias **14**, 115–124 (1991)
20. Lippman, S.A.: Semi-Markov decision processes with unbounded rewards. Management Science **19**, 717–731 (1973)
21. Liu, J.Y., Huang S.M.: Markov decision processes with distribution function criterion of first-passage time. Appl. Math. Optim. **43**, 187–201 (2001)
22. Liu, J.Y., Liu, K.: Markov decision programming—the first passage model with denumerable state space. Systems Sci. Math. Sci. **5**, 340–351 (1992)
23. Newell R. G. and Pizer W. A. Discounting the distant future: how much do uncertain rates increase valuation. J. Environ. Econ. Manage **46**, 52–71 (2003)
24. Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons. Inc., New York (1994)
25. Ross, S.M.: Average cost semi-Markov decision processes. J. Appl. Probab. **7**, 649–656 (1970)
26. Sack, B., Wieland, V.: Interest-rate smoothing and optimal monetary policy: a review of recent empirical evidence. J. Econ. Bus. **52**, 205–228 (2000)
27. Yong, J.M., Zhou, X.Y.: Stochastic Controls—Hamiltonian Systems and HJB Equations. Springer-Verlag, New York (1999)
28. Yu, S.X., Lin, Y.L., Yan, P.F.: Optimization models for the first arrival target distribution function in discrete time. J. Math. Anal. Appl. **225**, 193–223 (1998)
29. Zhang, L.L., Guo, X.P.: Constrained continuous-time Markov control processes with average criteria. Math. Meth. Oper. Res. **67**, 323–340 (2008)