应用概率统计 第 37 卷
第 4 期 2021 年 8 月

Chinese Journal of Applied Probability and Statistics
Aug., 2021, Vol. 37, No. 4, pp. 421-440
doi: 10.3969/j.issn.1001-4268.2021.04.007

The Existence of Optimal Control for Continuous-Time Markov Decision Processes in Random Environments

SHAO Jinghai ZHAO Kun*

(*Center for Applied Mathematics, Tianjin University, Tianjin, 300072, China*)

Abstract: In this work, we investigate the optimal control problem for continuous-time Markov decision processes with the random impact of the environment. We provide conditions to show the existence of optimal controls under finite-horizon criteria. These results are established by introducing some restriction on the regularity of the optimal controls and by developing a new compactification method for continuous-time Markov decision processes, which is originally used to solve the optimal control problem for diffusion processes.

Keywords: Markov decision process; finite-horizon criterion; regime-switching diffusions; relaxed control; randomized policy

2020 Mathematics Subject Classification: 90C40; 93E20; 60J27; 60K37

Citation: SHAO J H, ZHAO K. The existence of optimal control for continuous-time Markov decision processes in random environments [J]. Chinese J Appl Probab Statist, 2021, 37(4): 421–440.

§1. Introduction

Continuous-time Markov decision processes (CTMDPs) have been extensively studied and widely applied in various application fields such as telecommunication, queueing systems, population processes, epidemiology, and so on. See, for instance, the monographs [1, 2], the works [3–10] and references therein. As an illustrative example, we consider the controlled queueing systems. In a single-server queueing system, jobs or customers arrive, enter the queue, wait for service, receive service, and then leave the system. A decision-maker can control the system by deciding which jobs to be admitted to the queue, by increasing or decreasing the arrival rates or service rates in order to maximize the reward or minimize the cost of this system. There are many researches on CTMDPs under various optimality criteria. For example, the expected discounted, average and the finite-horizon optimality criteria have been well studied in [1, 2] and [5, 9, 11] amongst others.

*Corresponding author, E-mail: zhaokun@tju.edu.cn.
Received September 1, 2020. Revised December 7, 2020.

However, in realistic applications, the cost of raw materials or the price of products depends on not only the number of jobs or customers but also the prices of raw materials or products. In this work, we shall extend the classical CTMDPs to make these models more realistic by including the random effect of the market. A diffusion process on $\mathbb{R}^d$ is included to model the price process whose coefficients may be dependent on the continuous-time Markov chain. A decision-maker still controls the system by deciding the transition rate of the Markov chain, but the optimality criterion depends on both the diffusion process and the Markov chain. The coexistence of Markov chains and diffusion processes makes the optimality problem more difficult. The well developed methods in the study of CTMDPs such as in [1] and [3,5] do not work anymore. For instance, to deal with the infinite horizon expected discounted reward, it is quite crucial to establish the optimality equation based on the recursion approximation of the Laplace transform for the continuous-time Markov chain; see [1; Theorem 4.6] and [12; p. 121–122]. Nevertheless, the appearance of the second order differential operators associated with the diffusion process makes it harder to first establish the optimality equation and then to show the existence of the optimal control.

In this work, we develop a compactification method to provide some sufficient conditions on the existence of optimal controls. This kind of compactification method is usually used to study the optimal control problem for jump-diffusion processes, and has been well studied by many works including [8, 13–18]. See [16] for a complete list of references on the subject. In order to deal with CTMDPs in a random environment, we introduce ψ-relaxed controls as the class of admissible controls. The function ψ is used to characterize the regularity of the optimal controls. The class of ψ-relaxed controls contains all randomized stationary policies in some sense (see Section 2 for details). The randomized stationary policies have been extensively investigated in the study of CTMDPs; see for example the monograph [1]. The basic idea of our method is similar to that of Haussmann and Suo [16], but there are some essential differences on the measurability of the control policies. In [16], the controllers are assumed to have no information on the state of the studied system, so the admissible control policies are all adapted to some given σ-fields. However, to deal with CTMDPs, the control policy must be adapted to the σ-fields generated by the Markov chain in order to keep the Markovian property of the studied system. Therefore, the key difficulty of this work is to show that the jumping process remains to be a Markov chain under all admissible controls in current situation. Besides, concrete techniques raised in this work are also different to those in [16]. This can be reflected by the fact that this work can treat the terminal cost, however, [16] cannot (cf. [16; Remark

2.2]).

To be more precise, consider a Markov chain (Λ_t) on a denumerable state space $\mathcal{S}$ associated with the transition rate q-pair $(q(\theta, A; u), q(\theta; u))$, where $\theta \in \mathcal{S}$, $A \in \mathscr{B}(\mathcal{S})$, $u \in U$, and the action set U is a compact subset of $\mathbb{R}^k$. Let us consider further a diffusion process (X_t) satisfying the following stochastic differential equation (SDE):

$$\mathrm{d}X_t = b(X_t, \Lambda_t)\mathrm{d}t + \sigma(X_t, \Lambda_t)\mathrm{d}B_t, \tag{1}$$

where $b : \mathbb{R}^d \times \mathcal{S} \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathcal{S} \to \mathbb{R}^{d\times d}$, and (B_t) is a standard d-dimensional Brownian motion. The process (X_t) is used to model the price of raw materials or products, which is related not only to the randomness of the market characterized by the Brownian motion, but also to the number of jobs or the customers characterized by the Markov chain (Λ_t). Relaxed controls, known also as randomized policies, are considered in this paper. The following finite-horizon criterion is used:

$$\mathsf{E}\Big[\int_0^T f(t, X_t, \Lambda_t, \mu_t)\mathrm{d}t + g(X_T, \Lambda_T)\Big],$$

where $f : [0, T] \times \mathbb{R}^d \times \mathcal{S} \times U \to \mathbb{R}$ and $g : \mathbb{R}^d \times \mathcal{S} \to \mathbb{R}$ stand for the cost functions. Here and in the remainder of this paper, a measurable function $h : U \to \mathbb{R}$ is extended into a function on $\mathscr{P}(U)$, the collection of all probability measures on U, through:

$$h(\mu) := \int_U h(u)\mu(\mathrm{d}u), \qquad \mu \in \mathscr{P}(U),$$

whenever the integral is well defined.

Our contribution of this paper consists of two aspects: one is to include the random impact of the environment into the cost/reward function to provide more realistic models than classical CTMDPs in applications; another is to propose a new method to study the existence of optimal controls for CTMDPs, which generalizes the method of [15–18] in the setting of Markov chains. This method can also be generalized to deal with the history-dependent control problem investigated in [5], where the existence of optimal history-dependent control was left open.

This work is organized as follows: To focus on the development of compactification method in [16, 18] from the setting of diffusion processes to that of CTMDPs, we consider in Section 2 only the optimal control problem for classical CTMDPs without any random impact of the environment. In Section 3 we treat CTMDPs in a random environment, and show the existence of the optimal control under appropriate conditions.

§2. Optimal Markov Control for CTMDPs

In this part we aim to develop the compactification method in [16,18] from the setting of jump-diffusion processes to the setting of CTMDPs. To focus on this development and simplify the representation, we do not consider the impact of random environment in this section. We introduce the concept of ψ-relaxed control to ensure the Markovian property of the studied system, and discuss its connection with the classical randomized control policies studied, for instance, in [1,5,7]. In short, the class of ψ-relaxed controls is a subset of general randomized control policies in some sense, but contains all the randomized stationary policies and deterministic stationary policies. Randomized or deterministic stationary policies are two important kinds of policies having been extensively studied in [1–5,7,10] amongst others. In these works, many obtained optimal control policies are all stationary.

1) Formulation and Assumptions

Let $(\Omega, \mathscr{F}, \mathsf{P})$ be a probability space with the filtration $\{\mathscr{F}_t\}_{t\geqslant 0}$. $\{\mathscr{F}_t\}$ satisfies the usual condition, that is, $\mathscr{F}_t$ is right-continuous and $\mathscr{F}_0$ contains all the P-negligible events in $\mathscr{F}$. Let $\mathcal{S}$ be a countable state space. Let $U \subset \mathbb{R}^k$ be a compact set, and $\mathscr{P}(U)$ the collection of all probability measures over U. On $\mathscr{P}(U)$, define the L_1-Wasserstein distance between two probability measures μ and ν by:

$$W_1(\mu,\nu) = \inf\Big\{\int_{U\times U} |x-y|\pi(\mathrm{d}x,\mathrm{d}y);\ \pi\in\mathscr{C}(\mu,\nu)\Big\}, \tag{2}$$

where $\mathscr{C}(\mu,\nu)$ stands for the collection of all probability measures on $U\times U$ with marginal μ and ν respectively. Since U is compact, and hence is bounded, the weak topology of $\mathscr{P}(U)$ is equivalent to the topology induced by the L_1-Wasserstein distance. Also, this implies that $(\mathscr{P}(U), W_1)$ is a compact Polish space (cf. [19; Chapter 7]). We focus on the finite-horizon optimal control problem in this work, so let us fix a time $T>0$ throughout this work.

Let $\mathcal{S}$ be a denumerable state space endowed with discrete topology. Given $u\in U$, we call $(q(\theta;u), q(\theta,A;u))$ $(\theta\in\mathcal{S},\ A\in\mathscr{B}(\mathcal{S}))$ a *q-pair*, if for each $A\in\mathscr{B}(S)$, $\theta\mapsto q(\theta;u)$ and $\theta\mapsto q(\theta,A;u)$ are measurable; and for each $\theta\in\mathcal{S}$, $A\mapsto q(\theta,A;u)$ is a measure on $\mathcal{S}$, $q(\theta,\{\theta\};u)=0$, $q(\theta,\mathcal{S};u)\leqslant q(\theta;u)$. Moreover, it is called *conservative* if $q(\theta;u)=q(\theta,\mathcal{S};u)$ for all $\theta\in\mathcal{S}$. A function $h:\mathcal{S}\to[0,\infty)$ is called a compact function if for every $\alpha>0$, the set $\{\theta\in\mathcal{S}; h(\theta)\leqslant\alpha\}$ is compact.

In the following we collect the hypotheses used in this section:

(H1) $U \subset \mathbb{R}^k$ is a compact set for some $k \in \mathbb{N}$.

(H2) For each $u \in U$, $(q(\theta; u), q(\theta, A; u))$ is a conservative q-pair on $\mathcal{S}$. Moreover, $M := \sup\limits_{u \in U} \sup\limits_{\theta \in \mathcal{S}} q(\theta, \mathcal{S}; u) < \infty$.

(H3) For every $\theta \in \mathcal{S}$ and $A \in \mathscr{B}(\mathcal{S})$, the function $u \mapsto q(\theta, A; u)$ is continuous on U. For every $A \in \mathscr{B}(\mathcal{S})$, $u \in U$, the function $\theta \mapsto q(\theta, A; u)$ is continuous.

(H4) There exist a compact function $\Phi : \mathcal{S} \to [1, \infty)$, a compact set $B_0 \in \mathscr{B}(\mathcal{S})$, constants $\lambda > 0$ and $\kappa_0 < \infty$ such that

$$Q_u \Phi(\theta) := \int_{\mathcal{S}} q(\theta, \mathrm{d}\gamma; u)\Phi(\gamma) - q(\theta; u)\Phi(\theta) \leqslant \lambda \Phi(\theta) + \kappa_0 \mathbf{1}_{B_0}(\theta), \qquad \theta \in \mathcal{S},\ u \in U.$$

Remark 1 The boundedness of $q(\theta, \mathcal{S}; u)$ in (H2) ensures that the jumping process (Λ_t) owns almost surely finite number of jumping in every finite time interval. As an initiative investigation to include the random effect of the environment to the theory of CTMDPs, we impose simply the bounded condition (H2) of the transition rates. In the study of CTMDPs, there are some works to deal with unbounded transition rates. For example, in [5], the authors used a technique of approximations from bounded transition rates to unbounded ones to establish the existence of optimal Markovian controls. (H4) is called a drift condition, which is used to guarantee the non-explosion of the process (Λ_t) and to prove the tightness of the distributions of the Markov chains.

Let $\psi : [0, T] \to [0, \infty)$ be an increasing function such that

$$\lim_{r \to 0} \psi(r) = 0. \tag{3}$$

Consider the space $\mathcal{D}([0, T]; \mathscr{P}(U))$ of measurable maps from $[0, T]$ to the Polish space $(\mathscr{P}(U), W_1)$ that are right-continuous with left-limits. Endow $\mathcal{D}([0, T]; \mathscr{P}(U))$ with the Skorokhod topology, which makes $\mathcal{D}([0, T]; \mathscr{P}(U))$ a Polish space; see [20]. For $\mu : [0, T] \to \mathscr{P}(U)$ in $\mathcal{D}([0, T]; \mathscr{P}(U))$, put

$$w_\mu([a, b)) = \sup\{W_1(\mu_t, \mu_s);\ s, t \in [a, b)\}, \qquad a, b \in [0, T],\ a < b.$$

To describe compact sets in $\mathcal{D}([0, T]; \mathscr{P}(U))$, let us introduce the function

$$w''_\mu(\delta) = \sup \min\{W_1(\mu_t, \mu_{t_1}), W_1(\mu_t, \mu_{t_2})\}, \tag{4}$$

where the supremum is taken over t_1, t, and t_2 satisfying $t_1 \leqslant t \leqslant t_2$, $t_2 - t_1 \leqslant \delta$.

Definition 2 A ψ-relaxed control is a term $\alpha = (\Omega, \mathscr{F}, \mathscr{F}_t, \mathsf{P}, \Lambda_t, \mu_t, s, \theta)$ satisfying:

(i) $(s, \theta) \in [0, T] \times \mathcal{S}$;

(ii) $(\Omega, \mathscr{F}, \mathsf{P})$ is a probability space with the filtration $\{\mathscr{F}_t\}_{t\in[0,T]}$;

(iii) $\mu_t \in \mathscr{P}(U)$ is adapted to the σ-field generated by Λ_t, $t \mapsto \mu_t$ is in $\mathcal{D}([0,T]; \mathscr{P}(U))$ almost surely, and for every $\theta' \in \mathcal{S}$ the curve $t \mapsto \nu_t(\,\cdot, \theta') := \mu_t(\,\cdot \,|\, \Lambda_t = \theta')$ satisfies $w_\nu([t_1, t_2)) \leqslant \psi(t_2 - t_1)$, $0 \leqslant t_1 < t_2 \leqslant T$;

(iv) $(\Lambda_t)_{t\in[s,T]}$ is an $\mathscr{F}_t$-adapted, jumping process with $\Lambda_s = \theta$ and satisfies

$$\mathsf{P}(\Lambda_{t+\delta} \in A \,|\, \Lambda_t = \theta, \mu_t = \mu) - \mathbf{1}_A(\theta) = [q(\theta, A; \mu) - q(\theta; \mu)\mathbf{1}_A(\theta)]\delta + o(\delta) \tag{5}$$

provided $\delta > 0$.

The collection of all ψ-relaxed controls with initial value (s, θ) is denoted by $\widetilde{\Pi}_{s,\theta}$. The function ψ is used to characterize the regularity of the optimal controls.

The set $\widetilde{\Pi}_{s,\theta}$ consists of many interesting and well studied controls. We proceed to show that all the randomized stationary policies and deterministic stationary policies studied, for example, in [1, 3, 5, 7] are all associated with ψ-relaxed controls in a natural way.

Recall the definition of randomized Markov policies from [1]. A randomized Markov policy is a real-valued function $\pi_t(C \,|\, \theta')$ that satisfies the following conditions:

(i) For all $\theta' \in \mathcal{S}$ and $C \in \mathscr{B}(U)$, $t \mapsto \pi_t(C \,|\, \theta')$ is measurable on $[0, \infty)$.

(ii) For all $\theta' \in \mathcal{S}$ and $t \geqslant 0$, $C \mapsto \pi_t(C \,|\, \theta')$ is a probability measure on $\mathscr{B}(U)$, where $\pi_t(C \,|\, \theta')$ denotes the probability that an action in C is taken when the system's state is θ' at time t.

A randomized Markov policy $\pi_t(\mathrm{d}u \,|\, \theta')$ is said to be stationary if $\pi_t(\mathrm{d}u \,|\, \theta')$ is independent of time t.

For any ψ-relaxed control $\alpha = (\Omega, \mathscr{F}, \mathscr{F}_t, \mathsf{P}, \Lambda_t, \mu_t, s, \theta)$, we shall show that μ_t indeed acts as a randomized Markov policy $\pi_t(C \,|\, \theta)$. Firstly, since μ_t is adapted to the σ-field generated by Λ_t according to Definition 2, this yields that there exists a measurable map $F_t : \mathcal{S} \to \mathscr{P}(U)$ such that $\mu_t = F_t(\Lambda_t)$. (This is a result derived from the functional monotone class theorem in measure theory.) Thus, if $\Lambda_t = \theta'$ is given, then $\mu_t = F_t(\theta')$ is a fixed probability measure in $\mathscr{P}(U)$. We may rewrite μ_t as

$$\mu_t(\mathrm{d}u) = \sum_{\theta'\in\mathcal{S}} F_t(\theta')(\mathrm{d}u)\mathbf{1}_{\{\Lambda_t=\theta'\}}. \tag{6}$$

Condition (iii) of Definition 2 ensures that $F_t(\theta')$ is right-continuous with left-limits. So $\pi_t(\mathrm{d}u \,|\, \theta') := F_t(\theta')(\mathrm{d}u)$ satisfies the conditions (i) and (ii) of a randomized Markov policy.

Consequently, the class of ψ-relaxed controls is a subclass of randomized Markov policies in some sense.

Moreover, for a randomized stationary policy $\pi(\mathrm{d}u \,|\, \theta')$, let

$$\widetilde{\mu}_t = \sum_{\theta' \in \mathcal{S}} \pi(\mathrm{d}u \,|\, \theta') \mathbf{1}_{\Lambda_t = \theta'}, \qquad t \in [0, T]. \tag{7}$$

According to the path property of continuous-time Markov chains, it is clear that $(\widetilde{\mu}_t)$ defined by (7) satisfies the condition (iii) of Definition 2 with $\nu_t(\mathrm{d}u, \theta') = \pi(\mathrm{d}u \,|\, \theta')$ for all $t \geqslant 0$ and $\theta' \in \mathcal{S}$. Hence, $w_\nu([t_1, t_2)) = 0$ for every $0 \leqslant t_1 < t_2$. Corresponding to the randomized stationary Markov policy $\pi(\mathrm{d}u \,|\, \theta')$, there exists a CTMDPs (Λ_t) in some probability space $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathsf{P})$ with initial value $\Lambda_s = \theta$; see [1; Chapter 2]. It follows immediately that $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathsf{P}, \Lambda_t, \widetilde{\mu}_t, s, \theta)$ is a ψ-relaxed control for any ψ satisfying (3). By viewing a deterministic stationary policy $\xi : \mathcal{S} \to U$ as a randomized policy $\pi : \mathcal{S} \to \mathscr{P}(U)$ through the transform $\pi(\mathrm{d}u \,|\, \theta') = \mathbf{1}_{\xi(\theta')}(\mathrm{d}u)$, we know that every deterministic stationary policy is corresponding to a ψ-relaxed control.

Conditions (iii) and (iv) of Definition 2 also tell us that the transition rate does not depend on the past of the process (Λ_t), so the process (Λ_t) is indeed a Markov process. Put

$$q(t, \theta', A) = \mathsf{E}\Big[\int_U q(\theta', A; u)\mu_t(\mathrm{d}u) \,|\, \Lambda_t = \theta'\Big], \quad q(t, \theta') = \mathsf{E}\Big[\int_U q(\theta'; u)\mu_t(\mathrm{d}u) \,|\, \Lambda_t = \theta'\Big] \tag{8}$$

for $A \in \mathscr{B}(\mathcal{S})$, then the transition probability of the process (Λ_t) satisfies

$$\mathsf{P}(\Lambda_{t+\delta} \in A \,|\, \Lambda_t = \theta') - \mathbf{1}_A(\theta') = [q(t, \theta', A) - q(t, \theta')\mathbf{1}_A(\theta')]\delta + o(\delta). \tag{9}$$

Given two measurable functions $f : [0, T] \times \mathcal{S} \times U \to \mathbb{R}$ and $g : \mathcal{S} \to \mathbb{R}$, the expected cost under the policy $\alpha \in \widetilde{\Pi}_{s,\theta}$ is defined by

$$J(s, \theta, \alpha) = \mathsf{E}\Big[\int_s^T f(t, \Lambda_t, \mu_t)\mathrm{d}t + g(\Lambda_T)\Big], \qquad s \in [0, T),\ \theta \in \mathcal{S}. \tag{10}$$

Define the value function by

$$V(s, \theta) = \inf_{\alpha \in \widetilde{\Pi}_{s,\theta}} J(s, \theta, \alpha), \qquad s \in [0, T),\ \theta \in \mathcal{S}. \tag{11}$$

For $s \in [0, T]$, $\theta \in \mathcal{S}$, a ψ-relaxed control $\alpha^* \in \widetilde{\Pi}_{s,\theta}$ is called *optimal* if

$$V(s, \theta) = J(s, \theta, \alpha^*). \tag{12}$$

2) Existence of Optimal Control

After the preparation of the previous subsection, we can state our result on the existence of optimal ψ-relaxed controls. We shall follow Haussmann and Suo's approach, and one can refer to [5] for alternative approach in the setting of CTMDPs without the random impact of the environment.

Theorem 3 Given $T > 0$, assume (H1)−(H4) hold. Suppose f and g are lower semi-continuous and bounded from below. Then for every $s \in [0, T)$ and $\theta \in \mathcal{S}$ there exists an optimal ψ-relaxed control $\alpha^* \in \widetilde{\Pi}_{s,\theta}$.

Before proving this theorem, for a relaxed control $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathsf{P}, \Lambda_t, \mu_t, s, \theta)$ we provide a representation of the transition probability of the Markov chain (Λ_t). Define

$$P^{\mu}_{s,t}\mathbf{1}_A(\theta) = P^{\mu}(s, \theta, t, A) = \mathsf{P}(\Lambda_t \in A \,|\, \Lambda_s = \theta), \qquad \theta \in \mathcal{S},\ A \in \mathscr{B}(\mathcal{S}), \tag{13}$$

and

$$Q^{\mu}(t)h(\theta) = \int_{\mathcal{S}} q(t, \theta, \mathrm{d}\gamma)h(\gamma) - q(t, \theta)h(\theta), \qquad h \in \mathscr{B}_b(\mathcal{S}), \tag{14}$$

where $q(t, \theta, \cdot)$ and $q(t, \theta)$ are given by (8), $\mathscr{B}(\mathcal{S})$ denotes the set of measurable functions on $\mathcal{S}$, and $\mathscr{B}_b(\mathcal{S})$ is the set of bounded measurable functions on $\mathcal{S}$.

Proposition 4 For a relaxed control $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathsf{P}, \Lambda_t, \mu_t, s, \theta)$, it holds, for $h \in \mathscr{B}_b(\mathcal{S})$,

$$\begin{aligned} P^{\mu}_{s,t}h(\theta) = h(\theta) &+ \int_s^t Q^{\mu}(t_1)h(\theta)\mathrm{d}t_1 + \int_s^t \int_s^{t_2} Q^{\mu}(t_2)Q^{\mu}(t_1)h(\theta)\mathrm{d}t_1\mathrm{d}t_2 \\ &+ \sum_{n=3}^{\infty} \int_s^t \int_s^{t_n} \cdots \int_s^{t_2} Q^{\mu}(t_n)Q^{\mu}(t_{n-1}) \cdots Q^{\mu}(t_1)h(\theta)\mathrm{d}t_1 \cdots \mathrm{d}t_{n-1}\mathrm{d}t_n. \end{aligned} \tag{15}$$

Proof Due to (iv) of Definition 2 and (8), (9), we know that (Λ_t) is a time-inhomogeneous Markov process. Therefore,

$$P^{\mu}_{s,t+\delta}h(\theta) = P^{\mu}_{s,t}P^{\mu}_{t,t+\delta}h(\theta), \qquad h \in \mathscr{B}_b(\mathcal{S}).$$

Invoking (9), this yields the equation

$$\frac{\mathrm{d}}{\mathrm{d}t}P^{\mu}_{s,t}h(\theta') = P^{\mu}_{s,t}Q^{\mu}(t)h(\theta'), \quad P^{\mu}_{s,s}h(\theta') = h(\theta'), \qquad \theta' \in \mathcal{S},\ h \in \mathscr{B}_b(\mathcal{S}). \tag{16}$$

See, e.g. [21] for more details on this deduction. Thus, according to [22; Chapter III], formulae (1.12) and (1.15) therein, the unique solution of (16) has an explicit representation (15) in terms of the Cauchy operator.

Let us show the series in (15) is well defined. Endowed with the essential supremum norm $\|\cdot\|_\infty$, $\mathscr{B}_b(\mathcal{S})$ becomes a Banach space. Viewed as a linear operator over $\mathscr{B}_b(\mathcal{S})$, define the operator norm of $Q^\mu(t)$ by:

$$\|Q^\mu(t)\| = \sup_{\|h\|_\infty \leqslant 1} \|Q^\mu(t)h\|_\infty,$$

which obviously satisfies $\|Q^\mu(t)\| \leqslant \sup\limits_{\theta\in\mathcal{S}} \sup\limits_{u\in U} 2q(\theta;u) \leqslant 2M < \infty$, $\forall\, t \in [0,T]$. Hence,

$$\begin{aligned}
&\Big| \int_s^t \int_s^{t_n} \cdots \int_s^{t_2} Q^\mu(t_n)Q^\mu(t_{n-1}) \cdots Q^\mu(t_1)h(\theta)\mathrm{d}t_1 \cdots \mathrm{d}t_{n-1}\mathrm{d}t_n \Big| \\
&\leqslant \|h\|_\infty \int_s^t \int_s^{t_n} \cdots \int_s^{t_2} \|Q^\mu(t_n)\| \|Q^\mu(t_{n-1})\| \cdots \|Q^\mu(t_1)\| \mathrm{d}t_1 \cdots \mathrm{d}t_{n-1}\mathrm{d}t_n \\
&= \frac{\|h\|_\infty}{n!} \int_s^t \int_s^t \cdots \int_s^t \|Q^\mu(t_n)\| \|Q^\mu(t_{n-1})\| \cdots \|Q^\mu(t_1)\| \mathrm{d}t_1 \cdots \mathrm{d}t_{n-1}\mathrm{d}t_n \\
&= \frac{\|h\|_\infty}{n!} \Big[\int_s^t \|Q^\mu(r)\| \mathrm{d}r \Big]^n \leqslant \frac{[2M(t-s)]^n}{n!} \|h\|_\infty,
\end{aligned} \tag{17}$$

since the integral is invariant under any perturbation of the variables $t_1, t_2, \cdots, t_n$. Therefore, the series in (15) is convergent, and further the operator $P^\mu_{s,t}$ is well defined. □

Just as done in [16], the relaxed controls can be transformed into controls in the canonical path space to simplify the arguments. Let

$$\mathcal{U} = \{\nu : [0,T] \to \mathscr{P}(U);\, \nu \in \mathcal{D}([0,T];\mathscr{P}(U)),\, w''_\nu(\delta) \leqslant \psi(\delta),\, \delta \in (0,T]\}, \tag{18}$$

which is viewed as a subspace of $\mathcal{D}([0,T];\mathscr{P}(U))$. Denote

$$\mathcal{D}([0,T];\mathcal{S}) = \{y : [0,T] \to \mathcal{S} \text{ is right-continuous with left-limits}\},$$

which is a Polish space endowed with Skorokhod topology. Consider the canonical space $\mathcal{Y} = \mathcal{D}([0,T];\mathcal{S}) \times \mathcal{U}$. Let $\widetilde{\mathcal{D}}$, $\widetilde{\mathcal{U}}$ be their Borel σ-fields, and $\widetilde{\mathcal{D}}_t$, $\widetilde{\mathcal{U}}_t$ the σ-fields up to time t. Put $\widetilde{\mathcal{Y}} = \widetilde{\mathcal{D}} \times \widetilde{\mathcal{U}}$, $\widetilde{\mathcal{Y}}_t = \widetilde{\mathcal{D}}_t \times \widetilde{\mathcal{U}}_t$. Then, every ψ-relaxed control $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathsf{P}, \Lambda_t, \mu_t, s, \theta)$ can be transformed into a new ψ-relaxed control $(\mathcal{Y}, \widetilde{\mathcal{Y}}, \widetilde{\mathcal{Y}}_t, R, \Lambda_t, \mu_t, s, \theta)$ via the map $\Psi : \Omega \to \mathcal{Y}$ defined by

$$\Psi(\omega) = (\Lambda_t(\omega), \mu_t(\omega))_{t\in[0,T]}, \qquad \Lambda_r := \theta, \quad \mu_r := \mu_s, \quad \forall\, r \in [0,s],$$

where $R = \mathsf{P} \circ \Psi^{-1}$ is a probability measure on $\mathcal{Y}$. Similar to the discussion in [16], it is clear that the ψ-relaxed control $\alpha = (\mathcal{Y}, \widetilde{\mathcal{Y}}, \widetilde{\mathcal{Y}}_t, R, \Lambda_t, \mu_t, s, \theta)$ is completely determined by the probability measure R, so in the canonical space we use R itself to denote this ψ-relaxed control α.

Proof of Theorem 3 If $V(s,\theta)=\infty$, then every ψ-relaxed control α will be optimal. So, we only need to consider the case $V(s,\theta)<\infty$. We only consider the case $s=0$ to simplify the notation. The proof is separated into three steps.

Step 1: According to the definition of $V(0,\theta)$ and previously introduced representation of ψ-relaxed controls on the canonical space, there exists a sequence of probability measures R_n, $n\geqslant 1$, on $\mathcal{Y}$ such that

$$\lim_{n\to\infty} J(0,\theta,R_n)=V(0,\theta)<\infty. \tag{19}$$

In this step, we aim to prove that $(R_n)_{n\geqslant 1}$ is tight. To this end, let $\mathscr{L}_\Lambda^n$ and $\mathscr{L}_\mu^n$, $n\geqslant 1$, the marginal distribution of $(\Lambda_t)_{t\in[0,T]}$ and $(\mu_t)_{t\in[0,T]}$ respectively under R_n.

Since U is a compact set, $(\mathscr{P}(U),W_1)$ is a compact Polish space. Then, according to [20; Theorem 14.3] or [23; Theorem 6.3], $\mathcal{U}$ is a compact subset in $\mathcal{D}([0,T];\mathscr{P}(U))$. Moreover, by the definition of ψ-relaxed control, μ admits a representation (6), and $F_t(\theta')$ is in $\mathcal{U}$ for every $\theta'\in\mathcal{S}$. The compactness of $\mathscr{P}(U)$ implies the boundedness of $\mathscr{P}(U)$, i.e. there exists a constant $K>0$ such that $W_1(\nu_1,\nu_2)\leqslant K$ for any $\nu_1,\nu_2\in\mathscr{P}(U)$. This yields immediately that for some fixed $\nu\in\mathscr{P}(U)$,

$$R_n\Big(\omega:\sup_{0\leqslant t\leqslant T} W_1(\mu_t,\nu)>K\Big)=0,\qquad n\geqslant 1.$$

We go to estimate $R_n(\omega: w''_{\mu(\omega)}(\delta)\geqslant\varepsilon)$, $n\geqslant 1$. For any $\varepsilon\in(0,K)$, there exists a $\delta>0$ such that $\psi(\delta)<\varepsilon$. According to Definition 2, for every $\theta'\in\mathcal{S}$, denoting by $\nu_t(\cdot,\theta'):=\mu_t(\cdot\,|\,\Lambda_t=\theta')$, it holds

$$w_\nu([t_1,t_2))\leqslant\psi(t_2-t_1)\leqslant\psi(\delta)<\varepsilon,\qquad 0\leqslant t_1<t_2\leqslant T,\ t_2-t_1\leqslant\delta.$$

Also, we can rewrite $\mu_t(\cdot)=\nu_t(\cdot,\Lambda_t)$. By the triangle inequality,

$$\begin{aligned} W_1(\mu_t,\mu_{t_1}) &\leqslant W_1(\nu_t(\cdot,\Lambda_t),\nu_{t_1}(\cdot,\Lambda_t))+W_1(\nu_{t_1}(\cdot,\Lambda_t),\nu_{t_1}(\cdot,\Lambda_{t_1}))\\ &\leqslant W_1(\nu_t(\cdot,\Lambda_t),\nu_{t_1}(\cdot,\Lambda_t))+K\mathbf{1}_{\Lambda_t\neq\Lambda_{t_1}}. \end{aligned}$$

Hence, for any $t_1,t,t_2\in[0,T]$ with $t_1\leqslant t\leqslant t_2$ and $t_2-t_1\leqslant\delta$, if there exist no more than two jumps for the Markov chain (Λ_t) during the time period $[t_1,t_2]$, it must hold

$$\begin{aligned} &\min\{W_1(\mu_{t_1},\mu_t),W_1(\mu_{t_2},\mu_t)\}\\ \leqslant\ &\min\{W_1(\nu_t(\cdot,\Lambda_t),\nu_{t_1}(\cdot,\Lambda_t))+K\mathbf{1}_{\Lambda_t\neq\Lambda_{t_1}},W_1(\nu_t(\cdot,\Lambda_t),\nu_{t_2}(\cdot,\Lambda_t))+K\mathbf{1}_{\Lambda_t\neq\Lambda_{t_2}}\}<\varepsilon. \end{aligned}$$

Thus,

$$R_n(\omega:\min\{W_1(\mu_{t_1},\mu_t),W_1(\mu_{t_2},\mu_t)\}\geqslant\varepsilon)$$

$$\leqslant R_n(\omega : \text{the process } (\Lambda_r) \text{ owns at least two jumps during } [t_1, t_2]) \leqslant o(\delta). \tag{20}$$

Moreover, the arbitrariness of t_1, t, t_2 implies that for each positive ε and η, there exists $\delta \in (0, T)$ such that

$$R_n(\omega : w''_\mu(\delta) \geqslant \varepsilon) \leqslant o(\delta) \leqslant \eta. \tag{21}$$

For the Markov chain (Λ_t) with the bounded transition rate matrices, it is clear that for $\delta > 0$ sufficiently small,

$$R_n(\omega : w_\mu([0, \delta)) \geqslant \varepsilon) \leqslant \eta, \quad R_n(\omega : w_\mu([T - \delta, T)) \geqslant \varepsilon) \leqslant \eta, \qquad n \geqslant 1. \tag{22}$$

Applying [20; Theorem 15.3], we show that $(\mathscr{L}^n_\mu)_{n\geqslant 1}$ is tight.

Next, we go to prove the set of probability measures $(\mathscr{L}^n_\Lambda)_{n\geqslant 1}$ on $\mathcal{D}([0, T]; \mathcal{S})$ is tight. We shall apply Kurtz's tightness criterion (cf. [23; Theorem 8.6, p. 137]) to prove it.

On one hand, by (H4) and Itô's formula, we have

$$\mathsf{E}_{R_n}\Phi(\Lambda_t) = \Phi(\theta) + \mathsf{E}_{R_n}\int_0^t Q_{\mu_s}\Phi(\Lambda_s)\mathrm{d}s \leqslant \Phi(\theta) + \mathsf{E}_{R_n}\int_0^t [\lambda\Phi(\Lambda_s) + \kappa_0]\mathrm{d}s,$$

where E_{R_n} stands for taking expectation w.r.t. R_n. Then Gronwall's inequality leads to that

$$\mathsf{E}_{R_n}\Phi(\Lambda_t) \leqslant [\Phi(\theta) + \kappa_0 T]\mathsf{E}^{\lambda t}, \qquad t \in [0, T]. \tag{23}$$

Then, for any $\varepsilon > 0$, take N_ε large enough so that

$$\frac{\mathsf{E}_{R_n}\Phi(\Lambda_t)}{N_\varepsilon} \leqslant \frac{[\Phi(\theta) + \kappa_0 T]\mathsf{E}^{\lambda T}}{N_\varepsilon} < \varepsilon.$$

Let $K_\varepsilon = \{\gamma \in \mathcal{S}; \Phi(\gamma) \leqslant N_\varepsilon\}$, which is a compact set because Φ is a compact function. Then,

$$\sup_n R_n(\Lambda_t \in K^c_\varepsilon) \leqslant \sup_n \frac{\mathsf{E}_{R_n}\Phi(\Lambda_t)}{N_\varepsilon} < \varepsilon. \tag{24}$$

On the other hand, we also need to show that for any $\delta > 0$ there exists a nonnegative random variable $\gamma_n(\delta) \geqslant 0$ such that

$$\mathsf{E}_{R_n}[\mathbf{1}_{\Lambda_{t+u}\neq\Lambda_t} \,|\, \mathscr{F}_t] \leqslant \mathsf{E}_{R_n}[\gamma_n(\delta) \,|\, \mathscr{F}_t], \qquad 0 \leqslant t \leqslant T,\ 0 \leqslant u \leqslant \delta,$$

and $\lim\limits_{\delta\to 0}\sup\limits_n \mathsf{E}_{R_n}[\gamma_n(\delta)] = 0$. Under (H2), the transition rate $(q(\theta, A; u), q(\theta; u))$ of (Λ_t) is bounded, and hence

$$R_n(\Lambda_s = \Lambda_t,\ \forall\, s \in [t, t + u]) \geqslant \mathsf{E}_{R_n}\Big[\exp\Big(-\int_t^{t+u}\sup_{\theta\in\mathcal{S}} q(\theta; \mu_s)\mathrm{d}s\Big)\Big] \geqslant \exp(-Mu).$$

Then, for every $0 \leqslant u \leqslant \delta$,

$$\mathsf{E}_{R_n}[\mathbf{1}_{\{\Lambda_{t+u}\neq\Lambda_t\}}] \leqslant 1 - R_n(\Lambda_s = \Lambda_t,\, \forall\, s \in [t, t+u]) \leqslant 1 - \mathrm{e}^{-Mu} \leqslant 1 - \mathrm{e}^{-M\delta} =: \gamma_n(\delta).$$

It is clear that $\lim\limits_{\delta\to 0}\sup\limits_{n} \mathsf{E}_{R_n}\gamma_n(\delta) = 0$. Combining this with (24), we conclude that $(\mathscr{L}_\Lambda^n)_{n\geqslant 1}$ is tight.

As a consequence, the fact $(\mathscr{L}_\Lambda^n)_{n\geqslant 1}$ and $(\mathscr{L}_\mu^n)_{n\geqslant 1}$ are both tight leads to that for any $\varepsilon > 0$, there exist compact sets $K_1 \subset C([0,T];\mathscr{P}(U))$ and $K_2 \subset \mathcal{D}([0,T];\mathcal{S})$ such that

$$R_n(\mathcal{D}([0,T];\mathcal{S}) \times K_1^c) = \mathscr{L}_\mu^n(K_1^c) < \varepsilon, \qquad R_n(K_2^c \times \mathscr{P}([0,T]\times U)) = \mathscr{L}_\Lambda^n(K_2^c) < \varepsilon,$$

where K_i^c, $i = 1, 2$, stands for the complement of K_i. So,

$$R_n((K_1 \times K_2)^c) \leqslant R_n(\mathcal{D}([0,T];\mathcal{S}) \times K_1^c) + R_n(K_2^c \times \mathscr{P}([0,T]\times U)) < 2\varepsilon,$$

which implies the desired tightness of $(R_n)_{n\geqslant 1}$.

Step 2: We go to show the existence of the optimal ψ-relaxed control in this step. According to the result of Step 1, $(R_n)_{n\geqslant 1}$ is tight, and up to taking a subsequence, R_n converges weakly to some probability measure R_0 on $\mathcal{Y}$. According to Skorokhod's representation theorem (cf. [23; Chapter 3], Theorem 1.8, p. 102), there exists a probability space $(\Omega', \mathscr{F}', \mathsf{P}')$ on which are defined $\mathcal{Y}$-valued random variables $Y_n = (\Lambda_t^{(n)}, \mu_t^{(n)})_{t\in[0,T]}$, $n = 1, 2, \cdots$, and $Y_0 = (\Lambda_t^{(0)}, \mu_t^{(0)})_{t\in[0,T]}$ with distribution R_n, $n = 1, 2, \cdots$, and R_0 respectively such that

$$\lim_{n\to\infty} Y_n = Y_0, \qquad \mathsf{P}'\text{-a.s..} \tag{25}$$

Denote $\mathscr{F}_t'$ the natural σ-field generated by $(\Lambda_s^{(n)}, \mu_s^{(n)})$, $n = 0, 1, 2, \cdots$, up to time t. We shall prove that $\alpha^* = (\Omega', \mathscr{F}', \mathscr{F}_t', \mathsf{P}', \Lambda_t^{(0)}, \mu_t^{(0)}, 0, \theta)$ is an optimal ψ-relaxed control with respect to the value function $V(0,\theta)$. To this end, we need to check that α^* satisfies the conditions of Definition 2. Obviously, conditions (i) and (ii) of Definition 2 hold.

To check condition (iv), the transition semigroup of $(\Lambda_t^{(n)})$, $P_{s,t}^{\mu^{(n)}}\mathbf{1}_A(\theta') := \mathsf{P}'(\Lambda_t^{(n)} \in A \,|\, \Lambda_s^{(n)} = \theta')$, $\theta' \in \mathcal{S}$, $A \in \mathscr{B}(\mathcal{S})$, is determined by the equation (15) with $Q^\mu(t)$ being replaced by $Q^{\mu^{(n)}}(t)$ defined as follows:

$$\begin{aligned} Q^{\mu^{(n)}}(t)h(\theta') = {} & \mathsf{E}\Big[\int_U\int_{\mathcal{S}} q(\theta', \mathrm{d}\gamma; u)h(\gamma)\mu_t^{(n)}(\mathrm{d}u) \,\Big|\, \Lambda_t^{(n)} = \theta'\Big] \\ & - \mathsf{E}\Big[\int_U q(\theta'; u)\mu_t^{(n)}(\mathrm{d}u)h(\theta') \,\Big|\, \Lambda_t^{(n)} = \theta'\Big]. \end{aligned} \tag{26}$$

Similarly, we can define the operators $P_{s,t}^{\mu^{(0)}}$ and $Q^{\mu^{(0)}}(t)$.

For $0 \leqslant t_1 < t_2 < \cdots < t_k \leqslant T$, define the projection map $\pi_{t_1 t_2 \cdots t_k} : \mathcal{D}([0,T];\mathcal{S}) \to \mathcal{S}^k$ by $\pi_{t_1 t_2 \cdots t_k}(\Lambda_\cdot) = (\Lambda_{t_1}, \Lambda_{t_2}, \cdots, \Lambda_{t_k})$. Let $\mathcal{T}_0$ consist of those $t \in [0,T]$ for which the projection $\pi_t : \mathcal{D}([0,T];\mathcal{S}) \to \mathcal{S}$ is continuous except at points form a set of R_0-measure 0. For $t \in [0,T]$, $t \in \mathcal{T}_0$ if and only if $R_0(J_t) = 0$, where $J_t = \{\Lambda \in \mathcal{D}([0,T];\mathcal{S}); \Lambda_t \neq \Lambda_{t-}\}$. Also, 0, $T \in \mathcal{T}_0$ by convention. As a probability measure on $\mathcal{D}([0,T];\mathcal{S})$, it is known that the complement of $\mathcal{T}_0$ in $[0,T]$ is at most countable (cf. [20; p. 124]). Analogously, define the projection map $\widetilde{\pi}_{t_1 t_2 \cdots t_k} : \mathcal{U} \to \mathscr{P}(U)^k$ by $\widetilde{\pi}_{t_1 t_2 \cdots t_k}(\mu_\cdot) = (\mu_{t_1}, \mu_{t_2}, \cdots, \mu_{t_k})$, which is clearly continuous.

Since $(\Lambda_t^{(n)}, \mu_t^{(n)})_{t\in[0,T]}$ converges almost surely to $(\Lambda_t^{(0)}, \mu_t^{(0)})_{t\in[0,T]}$ in the product space $\mathcal{D}([0,T];\mathcal{S}) \times \mathcal{U}$ as $n \to \infty$ and $\pi_t \times \widetilde{\pi}_t$ is continuous for $t \in \mathcal{T}_0$, we obtain that $(\Lambda_t^{(n)}, \mu_t^{(n)})$ converges almost surely to $(\Lambda_t^{(0)}, \mu_t^{(0)})$ for $t \in \mathcal{T}_0$. Since $T \in \mathcal{T}_0$, this implies, in particular, that

$$\Lambda_T^{(n)} \text{ converges almost surely to } \Lambda_T^{(0)} \text{ as } n \to \infty. \tag{27}$$

Letting $n \to \infty$ in (26) for $t \in \mathcal{T}_0$, we obtain

$$\lim_{n\to\infty} Q^{\mu^{(n)}}(t)h(\theta') = Q^{\mu^{(0)}}(t)h(\theta'), \qquad h \in \mathscr{B}_b(\mathcal{S}),\ \theta' \in \mathcal{S}.$$

For $t \in \mathcal{T}_0$, it holds

$$\lim_{n\to\infty} \mathsf{P}'(\Lambda_t^{(n)} \in A \,|\, \Lambda_0^{(n)} = \theta) = \mathsf{P}'(\Lambda_t^{(0)} \in A \,|\, \Lambda_0^{(0)} = \theta), \qquad A \in \mathscr{B}(\mathcal{S}),\ \theta \in \mathcal{S}. \tag{28}$$

Moreover, according to [23; Theorem 7.8, p. 131], for every $t \in [0,T]$, there exists a sequence $\{s_n\}_{n\geqslant 1}$ decreasing to t and $\Lambda_{s_n}^{(n)}$ converges weakly to $\Lambda_t^{(0)}$. For every $t \in [0,T]$, letting $n \to \infty$ in the following equation

$$\begin{aligned} P_{0,s_n}^{\mu^{(n)}}h(\theta') = h(\theta') &+ \int_0^{s_n} Q^{\mu^{(n)}}(t_1)h(\theta')\mathrm{d}t_1 + \int_0^{s_n}\int_0^{t_2} Q^{\mu^{(n)}}(t_2)Q^{\mu^{(n)}}(t_1)h(\theta')\mathrm{d}t_1\mathrm{d}t_2 \\ &+ \sum_{k=3}^{\infty}\int_0^{s_n}\int_0^{t_k}\cdots\int_0^{t_2} Q^{\mu^{(n)}}(t_k)Q^{\mu^{(n)}}(t_{k-1})\cdots Q^{\mu^{(n)}}(t_1)h(\theta')\mathrm{d}t_1\cdots\mathrm{d}t_{k-1}\mathrm{d}t_k, \end{aligned} \tag{29}$$

we obtain that

$$\begin{aligned} P_{0,t}^{\mu^{(0)}}h(\theta') = h(\theta') &+ \int_0^{t} Q^{\mu^{(0)}}(t_1)h(\theta')\mathrm{d}t_1 + \int_0^{t}\int_0^{t_2} Q^{\mu^{(0)}}(t_2)Q^{\mu^{(0)}}(t_1)h(\theta')\mathrm{d}t_1\mathrm{d}t_2 \\ &+ \sum_{k=3}^{\infty}\int_0^{t}\int_0^{t_k}\cdots\int_0^{t_2} Q^{\mu^{(0)}}(t_k)Q^{\mu^{(0)}}(t_{k-1})\cdots Q^{\mu^{(0)}}(t_1)h(\theta')\mathrm{d}t_1\cdots\mathrm{d}t_{k-1}\mathrm{d}t_k. \end{aligned} \tag{30}$$

Because the right-hand side of (30) is continuous in t, we have from (30) that $t \mapsto P_{0,t}^{\mu^{(0)}}h(\theta')$ is continuous. Whence, (9), and equivalently (5), is satisfied by taking derivative w.r.t. t in both sides of (30) and taking $h(\theta') = \mathbf{1}_A(\theta')$ for $A \in \mathscr{B}(\mathcal{S})$. This means that $(\Lambda_t^{(0)})$

is a continuous-time Markov chain associated with $(\mu_t^{(0)})$. As a consequence, there is no $t \in (0,T]$ such that $R_0(J_t) > 0$, and hence $\mathcal{T}_0 = [0,T]$.

Now we go to check condition (iii). Since $(\mu_t^{(n)})_{t\in[0,T]}$ converges almost surely to $(\mu_t^{(0)})_{t\in[0,T]}$ in $\mathcal{D}([0,T];\mathscr{P}(U))$, we have for each $t \in [0,T]$, $\mu_t^{(n)}$ converges almost surely to $\mu_t^{(0)}$ since $\mathcal{T}_0$ associated with $(\mu_t^{(0)})_{t\in[0,T]}$ equals to $[0,T]$. We adopt the notation in the study of backward martingale to define the filtration with negative indices. Let $\mathscr{F}_{-n}^{\Lambda} = \overline{\sigma(\Lambda_t^{(m)}, m \geqslant n)}$, the completion of the σ-field generated by $\Lambda_t^{(m)}$, $m \geqslant n$. Then

$$\mathscr{F}_{-1}^{\Lambda} \supset \mathscr{F}_{-2}^{\Lambda} \supset \cdots \supset \mathscr{F}_{-n}^{\Lambda} \supset \mathscr{F}_{-n-1}^{\Lambda} \supset \cdots .$$

Put $\mathscr{F}_{-\infty}^{\Lambda} = \bigcap\limits_{n\geqslant 1} \mathscr{F}_{-n}^{\Lambda}$. $\mathscr{F}_{-\infty}^{\Lambda}$ is easily checked to be a σ-field which concerns only the limit behavior of the sequence $\Lambda_t^{(n)}$, $n \geqslant 1$. Moreover, since there is no point in $[0,T]$ such that $(\Lambda_t^{(0)})$ must jump at that point with positive probability. Therefore, $\lim\limits_{n\to\infty} \Lambda_t^{(n)} = \Lambda_t^{(0)}$ a.s. for every $t \in [0,T]$, and further $\mathscr{F}_{-\infty}^{\Lambda} = \overline{\sigma(\Lambda_t^{(0)})}$. Define $\mathscr{F}_{-n}^{\mu} = \overline{\sigma(\mu_t^{(m)}, m \geqslant n)}$. Due to Definition 2 (iii), $\mu_t^{(n)}$ is in $\mathscr{F}_{-n}^{\Lambda}$ for each $n \geqslant 1$, and hence $\mathscr{F}_{-n}^{\mu} \subset \mathscr{F}_{-n}^{\Lambda}$. Therefore, it follows from the fact $\lim\limits_{n\to\infty} W_1(\mu_t^{(n)}, \mu_t^{(0)}) = 0$ a.s. that

$$\overline{\sigma(\mu_t^{(0)})} \subset \bigcap_{n\geqslant 1} \mathscr{F}_{-n}^{\mu} \subset \mathscr{F}_{-\infty}^{\Lambda} = \overline{\sigma(\Lambda_t^{(0)})},$$

which means that $\mu_t^{(0)}$ is adapted to $\sigma(\Lambda_t^{(0)})$.

Step 3: Invoking (27), (25), (19), and (10), we obtain by the lower semi-continuity of f and g that

$$\begin{aligned} V(0,\theta) &= \lim_{n\to\infty} \mathsf{E}_{\mathsf{P}'}\Big[\int_0^T f(t,\Lambda_t^{(n)},\mu_t^{(n)})\mathrm{d}t + g(\Lambda_T^{(n)})\Big] \\ &= \lim_{n\to\infty} \mathsf{E}_{\mathsf{P}'}\Big[\int_0^T \int_U f(t,\Lambda_t^{(n)},u)\mu_t^{(n)}(\mathrm{d}u)\mathrm{d}t + g(\Lambda_T^{(n)})\Big] \\ &\geqslant \mathsf{E}_{\mathsf{P}'}\Big[\int_0^T \int_U f(t,\Lambda_t^{(0)},u)\mu_t^{(0)}(\mathrm{d}u)\mathrm{d}t + g(\Lambda_T^{(0)})\Big] \geqslant V(0,\theta). \end{aligned} \tag{31}$$

Hence, α^* is an optimal ψ-relaxed control. The proof of this theorem is completed. □

After the existence of optimal ψ-relaxed control has been established, it is easy to use the time shift technique to prove the continuous property of the value function $V(s,\theta)$ under suitable condition of the cost functions; Moreover, based on the Dynkin formula, we can get a lower bound of the value function as follows. Suppose there exists a measurable function $\varphi : [0,T] \times \mathcal{S} \to \mathbb{R}$ satisfying $t \mapsto \varphi(t,\theta)$ is differentiable and

$$\varphi'(t,\theta) + f(t,\theta,u) + \sum_{\ell\in\mathcal{S}} q(\theta,\{\ell\};u)\varphi(t,\ell) - q(\theta;u)\varphi(t,\theta) \geqslant 0, \qquad \varphi(T,\theta) = g(\theta),$$

for every $t \in [0,T]$, $\theta \in \mathcal{S}$, $u \in U$. Then

$$V(s,\theta) \geqslant \varphi(s,\theta), \qquad s \in [0,T],\ \theta \in \mathcal{S}.$$

See, for example, [5; Section 3] for more details.

§3. Optimal Markov Control for CTMDPs in a Random Environment

In this section, we consider the random impact of the environment to CTMDPs. In such situation, the cost function depends not only on the paths of continuous-time Markov chains, but also on a stochastic process used to characterize, for instance, the price of raw materials. Precisely, such a dynamical system consists of two components: a diffusion process (X_t) and a continuous-time Markov chain (Λ_t), which is also called a regime-switching diffusion process; see, [24] and [25] and references therein. The process (X_t) is determined by the following SDE:

$$\mathrm{d}X_t = b(X_t, \Lambda_t)\mathrm{d}t + \sigma(X_t, \Lambda_t)\mathrm{d}B_t, \tag{32}$$

where (B_t) is a Brownian motion in $\mathbb{R}^d$; (Λ_t) is a continuous-time Markov process on the state space $\mathcal{S}$ associated with the q-pair $(q(\theta;u),\ q(\theta,A;u))$ satisfying

$$\mathsf{P}(\Lambda_{t+\delta} \in A \,|\, \Lambda_t = \theta, \mu_t = \mu) - \mathbf{1}_A(\theta) = [q(\theta,A;\mu) - q(\theta;\mu)\mathbf{1}_A(\theta)]\delta + o(\delta) \tag{33}$$

provided $\delta > 0$. The decision-maker still tries to minimize the cost through controlling the transition rates of the Markov chain (Λ_t), but now the cost function may depend on the diffusion process (X_t). Such kind of control problem is quite different to the usual studied optimal controls for SDEs (see, e.g. [15, 16]) or optimal controls for SDEs with regime-switching (see, e.g. [26–28]), where the control policies are placed directly to the drifts or diffusion coefficients of (X_t). Namely, the controlled system is also given by

$$\mathrm{d}\widetilde{X}_t = b(\widetilde{X}_t, \mu_t)\mathrm{d}t + \sigma(\widetilde{X}_t, \mu_t)\mathrm{d}B_t. \tag{34}$$

Roughly speaking, for $(\widetilde{X}_t)$, if we change the value of the control μ_t at time t, then the speed of $\widetilde{X}_t$ is immediately modified. Nevertheless, for (X_t) given by (32), if we change μ_t at time t, we only change the switching rate of the process (Λ_t) and the speed of X_t maybe remain the same as before because Λ_t may not jump at t. This observation tells us that in contrast to the process $(\widetilde{X}_t)$, the process (X_t) characterized by (32) and (33) is more closely related to the long time behavior of the control (μ_t).

Let ψ, $w''_\mu(\delta)$ be defined by (3) and (4) respectively.

Definition 5 A ψ-relaxed control is a term $\alpha = (\Omega, \mathscr{F}, \mathscr{F}_t, \mathsf{P}, B_t, X_t, \Lambda_t, \mu_t, s, x, \theta)$ such that

(i) $(s, x, \theta) \in [0, T] \times \mathbb{R}^d \times \mathcal{S}$;

(ii) $(\Omega, \mathscr{F}, \mathsf{P})$ is a probability space with the filtration $\{\mathscr{F}_t\}_{t\in[0,T]}$;

(iii) (B_t) is a d-dimensional standard Brownian motion on $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathsf{P})$, and (X_t, Λ_t) is a stochastic process on $\mathbb{R}^d \times \mathcal{S}$ satisfying (32) and (33) with $X_s = x$, $\Lambda_s = \theta$;

(iv) $\mu_t \in \mathscr{P}(U)$ is adapted to the σ-field generated by Λ_t, $t \mapsto \mu_t$ is in $\mathcal{D}([0, T]; \mathscr{P}(U))$ almost surely, and for every $\theta' \in \mathcal{S}$ the curve $t \mapsto \nu_t(\,\cdot, \theta'\,) := \mu_t(\,\cdot \,|\, \Lambda_t = \theta')$ satisfies $w_\nu([t_1, t_2)) \leqslant \psi(t_2 - t_1)$, $0 \leqslant t_1 < t_2 \leqslant T$.

The collection of all ψ-relaxed controls with initial value (s, x, θ) is denoted by $\widetilde{\Pi}_{s,x,\theta}$.

Remark 6 In Definition 5 (iv), the control policy μ_t is assumed to be adapted to the σ-field generated by Λ_t in order to ensure the controlled process (Λ_t) remain to be a Markov chain. In realistic applications, one may make a decision using the information of X_t. In that case, we naturally need to assume μ_t is adapted to the σ-field generated by Λ_t and X_t. But, Λ_t is no longer a Markov process.

Given two functions $f : [0, T] \times \mathbb{R}^d \times \mathcal{S} \times U \to \mathbb{R}$ and $g : \mathbb{R}^d \times \mathcal{S} \to \mathbb{R}$, the expected cost relative to the control $\alpha \in \widetilde{\Pi}_{s,x,\theta}$ is defined by

$$J(s, x, \theta, \alpha) = \mathsf{E}\Big[\int_s^T f(t, X_t, \Lambda_t, \mu_t)\mathrm{d}t + g(X_T, \Lambda_T)\Big]. \tag{35}$$

Correspondingly, the value function is defined by

$$V(s, x, \theta) = \inf_{\alpha\in\widetilde{\Pi}_{s,x,\theta}} J(s, x, \theta, \alpha) \tag{36}$$

for $s \in [0, T]$, $x \in \mathbb{R}^d$, $\theta \in \mathcal{S}$. A ψ-relaxed control $\alpha^* \in \widetilde{\Pi}_{s,x,\theta}$ is called optimal, if it holds

$$V(s, x, \theta) = J(s, x, \theta, \alpha^*).$$

We assume that the coefficients of (32) satisfy the following conditions.

(H5) There exists a constant $C_1 > 0$ such that

$$|b(x, \theta) - b(y, \theta)|^2 + \|\sigma(x, \theta) - \sigma(y, \theta)\|^2 \leqslant C_1|x - y|^2, \qquad x, y \in \mathbb{R}^d,\ \theta \in \mathcal{S},$$

where $|x|^2 = \sum_{k=1}^d x_k^2$, $\|\sigma\|^2 = \mathrm{tr}(\sigma\sigma')$, and σ' is the transpose of the matrix σ.

(H6) There exists a constant $C_2 > 0$ such that

$$|b(x,\theta)|^2 + \|\sigma(x,\theta)\|^2 \leqslant C_2(1+|x|^2), \qquad x \in \mathbb{R}^d,\ \theta \in \mathcal{S}.$$

The conditions (H5) and (H6) are classical conditions to ensure the existence and uniqueness of nonexplosive solution of SDE (1). These conditions can be weakened to include some non-Lipschitz coefficients (cf. e.g. [29]) or singular coefficients (cf. e.g. [30]).

Our second main result of this work is the following theorem.

Theorem 7 Assume that (H1)−(H6) hold, and f and g are lower semi-continuous and bounded from below. Then for every $s \in [0,T)$, $x \in \mathbb{R}^d$, $\theta \in \mathcal{S}$, there exists an optimal ψ-relaxed control $\alpha^* \in \widetilde{\Pi}_{s,x,\theta}$.

To simplify the proof, we also transform the relaxed controls into the canonical path space. Let $\mathcal{U}$ be defined by (18), and

$$\mathcal{Y} = C([0,T];\mathbb{R}^d) \times \mathcal{D}([0,T];\mathcal{S}) \times \mathcal{U}, \tag{37}$$

endowed with the product topology. Let $\widetilde{\mathcal{Y}}$ be the Borel σ-field, $\widetilde{\mathcal{Y}}_t$ the σ-fields up to time t. Now, the relaxed control $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathsf{P}, B_t, X_t, \Lambda_t, \mu_t, s, x, \theta)$ can be transformed into a relaxed control in the canonical space $\mathcal{Y}$ via the map $\Psi : \Omega \to \mathcal{Y}$ defined by

$$\Psi(\omega) = (X_t(\omega), \Lambda_t(\omega), \mu_t(\omega))_{t\in[0,T]}, \qquad X_r := x, \quad \Lambda_r := \theta, \quad \mu_r := \mu_s, \quad \forall\, r \in [0,s],$$

where $R = \mathsf{P} \circ \Psi^{-1}$ is a probability measure on $\mathcal{Y}$. In this canonical space, we still use R to represent this relaxed control $(\mathcal{Y}, \widetilde{\mathcal{Y}}, \widetilde{\mathcal{Y}}_t, R, B_t, X_t, \Lambda_t, \mu_t, s, x, \theta)$.

Proof of Theorem 7 Without loss of generality, we consider the case $V(0,x,\theta) < \infty$. In the canonical space $\mathcal{Y}$, there exists a sequence of probability measures R_n, $n \geqslant 1$, such that

$$\lim_{n\to\infty} J(0,x,\theta,R_n) = V(0,x,\theta) < \infty. \tag{38}$$

Step 1: In this step, we aim to prove the tightness of $(R_n)_{n\geqslant 1}$. Denote by $\mathscr{L}_X^n$, $\mathscr{L}_\Lambda^n$ and $\mathscr{L}_\mu^n$, $n \geqslant 1$, the distribution of $(X_t)_{t\in[0,T]}$, $(\Lambda_t)_{t\in[0,T]}$ and $(\mu_t)_{t\in[0,T]}$ respectively under R_n.

In the same way as the proof of Theorem 3, we can prove the tightness of $(\mathscr{L}_\mu^n)_{n\geqslant 1}$ and $(\mathscr{L}_\Lambda^n)_{n\geqslant 1}$. Now, we go to prove the tightness of $(\mathscr{L}_X^n)$. According to [20; Theorem 12.3], it is sufficient to verify the moment condition. By Itô's formula, for $0 \leqslant t_1 < t_2 \leqslant T$,

$$\mathsf{E}_{R_n}|X_{t_2} - X_{t_1}|^4$$

$$\leqslant 8\mathsf{E}_{R_n}\Big|\int_{t_1}^{t_2} b(X_r,\Lambda_r)\mathrm{d}r\Big|^4 + 8\mathsf{E}_{R_n}\Big|\int_{t_1}^{t_2}\sigma(X_r,\Lambda_r)\mathrm{d}B_r\Big|^4$$
$$\leqslant 8(t_2-t_1)^3\mathsf{E}_{R_n}\int_{t_1}^{t_2}|b(X_r,\Lambda_r)|^4\mathrm{d}r + 288(t_2-t_1)\mathsf{E}_{R_n}\int_{t_1}^{t_2}\|\sigma(X_r,\Lambda_r)\|^4\mathrm{d}r$$
$$\leqslant C(t_2-t_1)\int_{t_1}^{t_2}(1+\mathsf{E}_{R_n}|X_r|^4)\mathrm{d}r. \tag{39}$$

The linear growth condition (H6) implies the existence of a constant C (independent of n) such that $\int_0^T \mathsf{E}_{R_n}|X_r|^4\mathrm{d}r \leqslant C$ (cf. [24; Theorem 3.20]). Furthermore, invoking the fact $X_0 = x$, we conclude that $(\mathscr{L}_X^n)_{n\geqslant 1}$ is tight due to [20; Theorem 12.3].

Step 2: Because the marginal distributions of R_n, $n \geqslant 1$ are all tight, we get R_n, $n \geqslant 1$ is tight as well. Up to taking a subsequence, we may assume that R_n weakly converges to some probability measure R_0. Since $\mathcal{Y}$ is a Polish space, we apply Skorokhod's representation theorem (cf. [23; Chapter 3], Theorem 1.8, p. 102) to obtain a probability space $(\Omega', \mathscr{F}', \mathsf{P}')$ on which defined a sequence of random variables $(X_t^{(n)}, \Lambda_t^{(n)}, \mu_t^{(n)})_{t\in[0,T]}$, $n \geqslant 0$, taking values in $\mathcal{Y}$ with the distribution R_n, $n \geqslant 0$, respectively, such that $(X_t^{(n)}, \Lambda_t^{(n)}, \mu_t^{(n)})_{t\in[0,T]}$ converges P'-almost surely to $(X_t^{(0)}, \Lambda_t^{(0)}, \mu_t^{(0)})_{t\in[0,T]}$ as $n \to \infty$.

Let $\mathcal{T}_0$ be defined in the same way as the argument of Theorem 3. For every $t \in \mathcal{T}_0$, we have $(X_t^{(n)}, \Lambda_t^{(n)}, \mu_t^{(n)})$ converges almost surely to $(X_t^{(0)}, \Lambda_t^{(0)}, \mu_t^{(0)})$. Analogous to the argument of Theorem 3, $(\Lambda_t^{(0)})$ is a continuous time Markov chain with transition rate operator induced from $(\mu_t^{(0)})$, which also implies that $\mathcal{T}_0 = [0,T]$. The fact that $\mu_t^{(0)}$ is adapted to $\sigma(\Lambda_t^{(0)})$ can be proved in the same way as the proof of Theorem 3.

We need to check that $(X_t^{(0)})$ satisfies SDE (32) under R_0 is associated with a ψ-relaxed control. Since $(X_t^{(n)})_{t\in[0,T]}$ are processes in the path space $C([0,T];\mathbb{R}^d)$, every projection map $\pi_t : C([0,T];\mathbb{R}^d) \to \mathbb{R}^d$, $\pi_t(X.) := X_t$, is continuous. Then, this yields that

$$X_t^{(n)} \text{ converges almost surely to } X_t^{(0)} \text{ for each } t \in [0,T] \text{ as } n \to \infty,$$

because $(X_t^{(n)}, \Lambda_t^{(n)}, \mu_t^{(n)})_{t\in[0,T]}$ converges almost surely to $(X_t^{(0)}, \Lambda_t^{(0)}, \mu_t^{(0)})_{t\in[0,T]}$. Furthermore, passing n to ∞ in the following integral equation:

$$X_t^{(n)} = x + \int_0^t b(X_s^{(n)}, \Lambda_s^{(n)})\mathrm{d}s + \int_0^t \sigma(X_s^{(n)}, \Lambda_s^{(n)})\mathrm{d}B_s, \tag{40}$$

we get

$$X_t^{(0)} = x + \int_0^t b(X_s^{(0)}, \Lambda_s^{(0)})\mathrm{d}s + \int_0^t \sigma(X_s^{(0)}, \Lambda_s^{(0)})\mathrm{d}B_s, \tag{41}$$

which means that $(X_t^{(0)})$ satisfies SDE (32).

Consequently, R_0 is a ψ-relaxed control. By (38) and the lower semi-continuity of f and g, we have

$$\begin{aligned} V(0,x,\theta) &= \lim_{n\to\infty} \mathsf{E}_{\mathsf{P}'}\Big[\int_0^T f(t,X_t^{(n)},\Lambda_t^{(n)},\mu_t^{(n)})\mathrm{d}t + g(X_T^{(n)},\Lambda_T^{(n)})\Big] \\ &\geqslant \mathsf{E}_{\mathsf{P}'}\Big[\int_0^T f(t,X_t^{(0)},\Lambda_t^{(0)},\mu_t^{(0)})\mathrm{d}t + g(X_T^{(0)},\Lambda_T^{(0)})\Big] \geqslant V(0,x,\theta). \end{aligned}$$

Hence, R_0 is an optimal ψ-relaxed control. The proof is complete. □

References

[1] GUO X P, HERNÁNDEZ-LERMA O. *Continuous-Time Markov Decision Processes: Theory and Applications* [M]. Berlin: Springer-Verlag, 2009.

[2] PRIETO-RUMEAU T, HERNÁNDEZ-LERMA O. *Selected Topics on Continuous-Time Controlled Markov Chains and Markov Games* [M]. London: Imperial College Press, 2012.

[3] GUO X P. Continuous-time Markov decision processes with discounted rewards: the case of Polish spaces [J]. *Math Oper Res*, 2007, **32(1)**: 73–87.

[4] GUO X P, HERNÁNDEZ-LERMA O. Continuous-time controlled Markov chains [J]. *Ann Appl Probab*, 2003, **13(1)**: 363–388.

[5] GUO X P, HUANG X X, HUANG Y H. Finite-horizon optimality for continuous-time Markov decision processes with unbounded transition rates [J]. *Adv Appl Probab*, 2015, **47(4)**: 1064–1087.

[6] GUO X P, RIEDER U. Average optimality for continuous-time Markov decision processes in Polish spaces [J]. *Ann Appl Probab*, 2006, **16(2)**: 730–756.

[7] GUO X P, VYKERTAS M, ZHANG Y. Absorbing continuous-time Markov decision processes with total cost criteria [J]. *Adv Appl Probab*, 2013, **45(2)**: 490–519.

[8] KARATZAS I, SHREVE S E. Connections between optimal stopping and singular stochastic control I: monotone follower problems [J]. *SIAM J Control Optim*, 1984, **22(6)**: 856–877.

[9] MILLER B L. Finite state continuous time Markov decision processes with an infinite planning horizon [J]. *J Math Anal Appl*, 1968, **22(3)**: 552–569.

[10] PIUNOVSKIY A, ZHANG Y. Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach [J]. *SIAM J Control Optim*, 2011, **49(5)**: 2032–2061.

[11] YUSHKEVICH A A. Controlled Markov models with countable state space and continuous time [J]. *Theory Probab Appl*, 1978, **22(2)**: 215–235.

[12] ANDERSON W J. *Continuous-Time Markov Chains* [M]. New York: Springer, 1991.

[13] CHOW P L, MENALDI J L, ROBIN M. Additive control of stochastic linear systems with finite horizon [J]. *SIAM J Control Optim*, 1985, **23(6)**: 858–899.

[14] DUFOUR F, MILLER B. Maximum principle for singular stochastic control problems [J]. *SIAM J Control Optim*, 2006, **45(2)**: 668–698.

[15] HAUSSMANN U G, LEPELTIER J P. On the existence of optimal controls [J]. *SIAM J Control Optim*, 1990, **28(4)**: 851–902.

[16] HAUSSMANN U G, SUO W L. Singular optimal stochastic controls I: existence [J]. *SIAM J Control Optim*, 1995, **33(3)**: 916–936.

[17] HAUSSMANN U G, SUO W L. Singular optimal stochastic controls II: dynamic programming [J]. *SIAM J Control Optim*, 1995, **33(3)**: 937–959.

[18] KUSHNER H J. Existence results for optimal stochastic controls [J]. *J Optim Theory Appl*, 1975, **15(4)**: 347–359.

[19] AMBROSIO L, GIGLI N, SAVARÉ G. *Gradient Flows: In Metric Spaces and in the Space of Probability Measures* (*Lectures in Mathematics ETH Zürich*) [M]. Basel: Birkhäuser Verlag, 2005.

[20] BILLINGSLEY P. *Convergence of Probability Measures* [M]. New York: Wiley, 1968.

[21] FELLER W. On the integro-differential equations of purely discontinuous Markoff processes [J]. *Trans Amer Math Soc*, 1940, **48(3)**: 488–515.

[22] DALECKIĬ J L, KREĬN M G. *Stability of Solutions of Differential Equations in Banach Space* (*Translations of Mathematical Monographs, Vol. 43*) [M]. Providence, RI: American Mathematical Society, 1974.

[23] ETHIER S N, KURTZ T G. *Markov Processes: Characterization and Convergence*, New York: Wiley, 1986.

[24] MAO X R, YUAN C G. *Stochastic Differential Equations with Markovian Switching* [M]. London: Imperial College Press, 2006.

[25] YIN G G, ZHU C. *Hybrid Switching Diffusions: Properties and Applications* (*Stochastic Modelling and Applied Probability, 63*) [M]. New York: Springer, 2010.

[26] SONG Q S, STOCKBRIDGE R H, ZHU C. On optimal harvesting problems in random environments [J]. *SIAM J Control Optim*, 2011, **49(2)**: 859–889.

[27] SONG Q S, ZHU C. On singular control problems with state constraints and regime-switching: a viscosity solution approach [J]. *Automatica J IFAC*, 2016, **70**: 66–73.

[28] ZHOU X Y, YIN G. Markowitz's mean-variance portfolio selection with regime switching: a continuous-time model [J]. *SIAM J Control Optim*, 2003, **42(4)**: 1466–1482.

[29] SHAO J H. Strong solutions and strong Feller properties for regime-switching diffusion processes in an infinite state space [J]. *SIAM J Control Optim*, 2015, **53(4)**: 2462–2479.

[30] KRYLOV N V, RÖCKNER M. Strong solutions of stochastic equations with singular time dependent drift [J]. *Probab Theory Related Fields*, 2005, **131(2)**: 154–196.

随机环境下连续时间马氏决策过程最优控制存在性

邵井海　　赵　坤

(天津大学应用数学中心, 天津, 300072)

摘　要: 本文研究随机环境对于连续时间马氏决策过程最优控制问题的影响, 给出有限水平最优控制存在的判别条件, 将研究扩散过程最优控制问题常用的紧致化方法推广到对连续时间马氏决策过程的研究.

关键词: 马氏决策过程; 有限水平准则; 带切换的扩散过程; 松弛控制; 随机策略

中图分类号: O211.62