The existence of optimal control for continuous-time Markov decision processes in random environments

Jinghai Shao† and Kun Zhao ‡

June 21, 2019

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. Under appropriate conditions, the value function is continuous and satisfies the dynamic programming principle. 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 jump-diffusion processes.

AMS subject Classification (2010): 90C40, 93E20, 60J27, 60K37

Keywords: Markov decision process, finite-horizon criterion, Regime-switching diffusions, relaxed control, randomized policy

*Supported in part by NNSFs of China (No. 11771327, 11431014, 11831014)
†Center for Applied Mathematics, Tianjin University, Tianjin 300072, China. Email: shaojh@tju.edu.cn.
‡Center for Applied Mathematics, Tianjin University, Tianjin 300072, China.
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 [12, 26], the works [10, 11, 13, 14, 15, 19, 24, 25] 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 [12, 26] and [13, 24, 32] amongst others.

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 control 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 [12] and [10, 13] do not work anymore. The appearance of the second order differential operators associated with the diffusion process makes it harder to first establish the optimality equation and show the existence of its solution with some necessary regularity, then to show further 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 was usually used to study the optimal control problem for jump-diffusion processes, and has been well studied by many works including [4, 7, 16, 17, 18, 19, 21]. See [17] for a complete list of references on the subject. In order to deal with CTMDPs in a random environment, we introduce $\psi$-relaxed controls as the class of admissible controls. The function $\psi$ is used to characterize the regularity of the optimal controls. The class of $\psi$-relaxed controls contains all randomized stationary policies in some sense (see Section
The randomized stationary policies have been extensively investigated in the study of CTMDPs; see for example the monograph [12]. The basic idea of our method is similar to that of Haussmann and Suo [17], but there is some essential difference on the measurability of the control policies. In [17], the controllers are assumed to have no information on the state of the studies system, so the admissible control policies are all adapted to some given \( \sigma \)-fields. However, to deal with CTMDPs, the control policy must be adapted to the \( \sigma \)-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 remain 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 [17]. This can be reflected by the fact that this work can treat the terminal cost, however, [17] cannot (cf. [17, Remark 2.2]).

The CTMDPs in a random environment are closely related to the stochastic processes with regime-switching. Recently, stochastic models with regime-switching are widely applied in mathematical finance, ecological and biological systems, engineer, etc.; see, for example, [2, 6, 22, 31] amongst others. In view of its wide application, this optimal control problems for regime-switching processes have been studied in the literature. For instance, [33] proposed a continuous-time Markowitz’s mean-variance portfolio selection model with regime-switching and obtained the efficient portfolio that minimizes the risk of terminal wealth when given a fixed expected terminal wealth. [28] and [29] studied the singular control problem for regime-switching processes and characterized the corresponding value function as a viscosity solution of certain system of coupled non-linear quasi-variational inequalities. Nevertheless, the control mechanism in this work is quite different to those in [33, 28, 29]. In [33, 28, 29], the controls are imposed directly on the dynamical system itself, but here we impose the control on the transition rates of the Markov chain, which can be interpreted as changing the switching rates of the random environment. The methods used in [33, 28, 29] cannot be used to the optimal control problem considered in this work. Hence, this work is also a new development in the study of the optimal control problem for stochastic processes with regime-switching.

To be more precise, consider a Markov chain \((\Lambda_t)\) on a denumerable state space \(S\) associated with the transition rate \(q\)-pair \((q(\theta, A; u), q(\theta; u))\), where \(\theta \in S, A \in \mathcal{B}(S)\), \(u \in U\), 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):

\[
dX_t = b(X_t, \Lambda_t)dt + \sigma(X_t, \Lambda_t)dB_t, \tag{1.1}
\]
where \( b : \mathbb{R}^d \times S \to \mathbb{R}^d \), \( \sigma : \mathbb{R}^d \times S \to \mathbb{R}^{d \times d} \), 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 \((\Lambda_t)\). Relaxed controls, known also as randomized policies, are considered in this paper. The following finite-horizon criterion is used:

\[
\mathbb{E}\left[ \int_0^T f(t, X_t, \Lambda_t, \mu_t) \, dt + g(X_T, \Lambda_T) \right],
\]

where \( f : [0, T] \times \mathbb{R}^d \times S \times U \to \mathbb{R} \) and \( g : \mathbb{R}^d \times S \to \mathbb{R} \) 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 \( \mathcal{P}(U) \), the collection of all probability measures on \( U \), through:

\[
h(\mu) := \int_U h(u) \mu(du), \quad \mu \in \mathcal{P}(U),
\]

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 [16, 17, 18, 21] in the setting of Markov chains. Moreover, the concept of \( \psi \)-relaxed control proposed in this work is of interest by itself, which is closely related to the well studied randomized policy (cf. e.g. [12]). See the subsection 2.1 below for the details.

This work is organized as follows: To focus on the development of compactification method in [17, 21] for the optimal control problem from the setting of diffusion processes to that of CTMDPs, we consider in Section 2 the optimal control problem for classical CTMDPs without any random impact of the environment. The class of \( \psi \)-relaxed controls is a subset of classical admissible controls studied, for instance, in [12, 13, 15]. Therefore, our existence result of optimal \( \psi \)-relaxed control is a little stronger than the existence of classical admissible controls. In Section 3 we treat CTMDPs in a random environment, and show the existence of the optimal control under appropriate conditions. Based on the existence of optimal control, we further establish the dynamic programming principle after showing the continuity of the value function in Section 4.
2 Optimal Markov control for CTMDPs

In this part we aim to develop the compactification method in [17, 21] 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 random impact of the market on the Markov chains in this section. We introduce the concept of $\psi$-relaxed control to ensure the Markovian property of the studied system, and discuss its connection with the classical admissible controls studied, for instance, in [12, 13, 15]. In short, the class of $\psi$-relaxed controls is a subset of classical admissible controls but contains all the randomized stationary policies and deterministic stationary policies in some sense. These two kinds of policies have been extensively studied in [10, 11, 12, 13, 15, 25, 26] amongst others.

2.1 Formulation and Assumptions

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

$$W_1(\mu, \nu) = \inf \left\{ \int_{U \times U} |x - y| \pi(dx, dy); \pi \in \mathcal{C}(\mu, \nu) \right\},$$  \hspace{1cm} (2.1)

where $\mathcal{C}(\mu, \nu)$ stands for the collection of all probability measures on $U \times U$ with marginal $\mu$ and $\nu$ respectively. Since $U$ is compact, and hence is bounded, the weak topology of $\mathcal{P}(U)$ is equivalent to the topology induced by the $L_1$-Wasserstein distance. Also, this implies that $(\mathcal{P}(U), W_1)$ is a compact Polish space (cf. [1, 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 $(q(\theta; u), q(\theta, A; u))$ $(\theta \in \mathcal{S}, A \in \mathcal{B}(\mathcal{S}))$ a $q$-pair, if for each $A \in \mathcal{B}(\mathcal{S}), \theta \mapsto q(\theta; u)$ 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 $\mathcal{S}$, $q(\theta, \{\theta\}; u) = 0$, $q(\theta, \mathcal{S}; u) \leq q(\theta; u)$. 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 $\{\theta \in \mathcal{S}; h(\theta) \leq \alpha\}$ 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 $S$. Moreover, $M := \sup_{u \in U} \sup_{\theta \in S} q(\theta, S; u) < \infty$.

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

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

$$Q_u \Phi(\theta) := \int_S q(\theta, d\gamma; u)\Phi(\gamma) - q(\theta; u)\Phi(\theta) \leq \lambda \Phi(\theta) + \kappa_0 1_{B_0}(\theta), \quad \theta \in S, \ u \in U.$$ 

Remark 2.1 The boundedness of $q(\theta, S; u)$ in (H2) ensures that the jumping process $(\Lambda_t)$ 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 [13], 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 $(\Lambda_t)$ 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

$$\lim_{r \to 0} \psi(r) = 0. \quad (2.2)$$

Consider the space $\mathcal{D}([0, T]; \mathcal{P}(U))$ of measurable maps from $[0, T]$ to the Polish space $\mathcal{P}(U)$ that are right-continuous with left-limits. Endow $\mathcal{D}([0, T]; \mathcal{P}(U))$ with the Skorokhod topology, which makes $\mathcal{D}([0, T]; \mathcal{P}(U))$ a Polish space; see [3]. For $\mu : [0, T] \to \mathcal{P}(U)$ in $\mathcal{D}([0, T]; \mathcal{P}(U))$, put

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

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

$$w_\mu''(\delta) = \sup \min \{W_1(\mu_t, \mu_{t_1}), W_1(\mu_t, \mu_{t_2})\}, \quad (2.3)$$

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

$$\lim_{r \to 0} \psi(r) = 0. \quad (2.2)$$

Consider the space $\mathcal{D}([0, T]; \mathcal{P}(U))$ of measurable maps from $[0, T]$ to the Polish space $\mathcal{P}(U)$ that are right-continuous with left-limits. Endow $\mathcal{D}([0, T]; \mathcal{P}(U))$ with the Skorokhod topology, which makes $\mathcal{D}([0, T]; \mathcal{P}(U))$ a Polish space; see [3]. For $\mu : [0, T] \to \mathcal{P}(U)$ in $\mathcal{D}([0, T]; \mathcal{P}(U))$, put

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

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

$$w_\mu''(\delta) = \sup \min \{W_1(\mu_t, \mu_{t_1}), W_1(\mu_t, \mu_{t_2})\}, \quad (2.3)$$

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

$$\lim_{r \to 0} \psi(r) = 0. \quad (2.2)$$

Consider the space $\mathcal{D}([0, T]; \mathcal{P}(U))$ of measurable maps from $[0, T]$ to the Polish space $\mathcal{P}(U)$ that are right-continuous with left-limits. Endow $\mathcal{D}([0, T]; \mathcal{P}(U))$ with the Skorokhod topology, which makes $\mathcal{D}([0, T]; \mathcal{P}(U))$ a Polish space; see [3]. For $\mu : [0, T] \to \mathcal{P}(U)$ in $\mathcal{D}([0, T]; \mathcal{P}(U))$, put

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

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

$$w_\mu''(\delta) = \sup \min \{W_1(\mu_t, \mu_{t_1}), W_1(\mu_t, \mu_{t_2})\}, \quad (2.3)$$
where the supremum is taken over \( t_1, t, \) and \( t_2 \) satisfying
\[
t_1 \leq t \leq t_2, \quad t_2 - t_1 \leq \delta.
\]

**Definition 2.2** A \( \psi \)-relaxed control is a term \( \alpha = (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, \Lambda_t, \mu_t, s, \theta) \) satisfying:

1. \((s, \theta) \in [0, T] \times S\);
2. \((\Omega, \mathcal{F}, \mathbb{P})\) is a probability space with the filtration \( \{\mathcal{F}_t\}_{t \in [0, T]}\);
3. \(\mu_t \in \mathcal{P}(U)\) is adapted to the \( \sigma \)-field generated by \( \Lambda_t \); for every \( t \geq 0 \), the curve \( t \mapsto \mu_t(\cdot | \Lambda_t = \theta') \) satisfies
   \[
   w_{\nu}(t_1, t_2) \leq \psi(t_2 - t_1), \quad 0 \leq t_1 < t_2 \leq T;
   \]
4. \((\Lambda_t)_{t \in [s, T]}\) is an \( \mathcal{F}_t \)-adapted, jumping process with \( \Lambda_s = \theta \) and satisfies
   \[
   \mathbb{P}(\Lambda_{t+\delta} \in A | \Lambda_t = \theta, \mu_t = \mu) - 1_A(\theta) = (q(\theta, A; \mu) - q(\theta; \mu)1_A(\theta))\delta + o(\delta) \quad (2.4)
   \]
   provided \( \delta > 0 \).

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

The set \( \mathcal{\tilde{\Pi}}_{s, \theta} \) 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 [10, 12, 13, 15] are all associated with \( \psi \)-relaxed controls in a natural way.

Recall the definition of randomized Markov policies from [12]. A randomized Markov policy is a real-valued function \( \pi_t(C|\theta) \) that satisfies the following conditions:

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

(ii) For all \( \theta' \in S \) and \( t \geq 0 \), \( C \mapsto \pi_t(C|\theta') \) is a probability measure on \( \mathcal{B}(U) \), where \( \pi_t(C|\theta') \) denotes the probability that an action in \( C \) is taken when the system’s state is \( \theta' \) at time \( t \).
A randomized Markov policy \( \pi_t(du|\theta') \) is said to be stationary if \( \pi_t(du|\theta') \) is independent of \( t \).

For any \( \psi \)-relaxed control \( \alpha = (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, \Lambda_t, \mu_t, s, \theta) \), we shall show that \( \mu_t \) indeed acts as a randomized Markov policy \( \pi_t(C|\theta) \). Firstly, since \( \mu_t \) is adapted to the \( \sigma \)-field generated by \( \Lambda_t \) according to Definition 2.2, this yields that there exists a measurable map \( F_t : S \to \mathcal{P}(U) \) such that \( \mu_t = F_t(\Lambda_t) \). Thus, if \( \Lambda_t = \theta' \) is given, then \( \mu_t = F_t(\theta') \) is a fixed probability measure in \( \mathcal{P}(U) \). We may rewrite \( \mu_t \) as

\[
\mu_t(du) = \sum_{\theta' \in S} F_t(\theta')(du) \mathbf{1}_{\{\Lambda_t = \theta'\}}.
\] (2.5)

Condition (3) of Definition 2.2 ensures that \( F_t(\theta') \) is right-continuous with left-limits. So \( \pi_t(du|\theta') := F_t(\theta')(du) \) satisfies the conditions (i) and (ii) of a randomized Markov policy. Consequently, the class of \( \psi \)-relaxed controls is a subclass of randomized Markov policies in some sense.

Moreover, for a randomized stationary policy \( \pi(du|\theta') \), let

\[
\tilde{\mu}_t = \sum_{\theta' \in S} \pi(du|\theta') \mathbf{1}_{\{\Lambda_t = \theta'\}}, \quad t \in [0, T].
\] (2.6)

According to the path property of continuous-time Markov chains, it is clear that \( (\tilde{\mu}_t) \) defined by (2.6) satisfies the condition (3) of Definition 2.2 with \( \nu_t(du, \theta') = \pi(du|\theta') \) for all \( t \geq 0 \) and \( \theta' \in S \). Hence, \( W_\nu([t_1, t_2]) = 0 \) for every \( 0 \leq t_1 < t_2 \). Corresponding to the randomized stationary Markov policy \( \pi(du|\theta') \), there exists a CTMDPs \( (\Lambda_t) \) in some probability space \( (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}) \) with initial value \( \Lambda_s = \theta \); see [12, Chapter 2]. It follows immediately that \( (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, \Lambda_t, \tilde{\mu}_t, s, \theta) \) is a \( \psi \)-relaxed control for any \( \psi \) satisfying (2.2). By viewing a deterministic stationary policy \( \xi : S \to U \) as a randomized policy \( \pi : S \to \mathcal{P}(U) \) through the transform

\[
\pi(du|\theta') = \mathbf{1}_{\xi(\theta')}(du),
\]

we know that every deterministic stationary policy is corresponding to a \( \psi \)-relaxed control.

Conditions (3) and (4) of Definition 2.2 also tell us that the transition rate does not depend on the past of the process \( \Lambda_t \), so the process \( \Lambda_t \) is indeed a Markov process. Put

\[
q(t, \theta', A) = \mathbb{E} \left[ \int_U q(\theta', A; u) \mu_t(du) | \Lambda_t = \theta' \right], \quad q(t, \theta') = \mathbb{E} \left[ \int_U q(\theta'; u) \mu_t(du) | \Lambda_t = \theta' \right]
\] (2.7)
for $A \in \mathcal{B}(\mathcal{S})$, then the transition probability of the process $(\Lambda_t)$ satisfies

$$P(\Lambda_{t+\delta} \in A|\Lambda_t = \theta') - 1_A(\theta') = (q(t, \theta', A) - q(t, \theta')1_A(\theta'))\delta + o(\delta). \quad (2.8)$$

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 $\tilde{\mu} \in \tilde{\Pi}$ is defined by

$$J(s, \theta, \alpha) = \mathbb{E} \left[ \int_s^T f(t, \Lambda_t, \mu_t)dt + g(\Lambda_T) \right], \quad s \in [0, T), \theta \in \mathcal{S}. \quad (2.9)$$

Define the value function by

$$V(s, \theta) = \inf_{\alpha \in \tilde{\Pi}_{s, \theta}} J(s, \theta, \alpha), \quad s \in [0, T), \theta \in \mathcal{S}. \quad (2.10)$$

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

$$V(s, \theta) = J(s, \theta, \alpha^*). \quad (2.11)$$

### 2.2 Existence of optimal control

After the preparation of the previous subsection, we can state our result on the existence of optimal $\psi$-relaxed controls. We shall follow Haussmann and Suo’s approach, and one can refer to [13] for alternative approach in the setting of CTMDPs without the random impact of the environment.

**Theorem 2.3** 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 $\psi$-relaxed control $\alpha^* \in \tilde{\Pi}_{s, \theta}$.

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

$$P_{s,t}^\mu 1_A(\theta) = P^\mu(s, \theta, t, A) = \mathbb{P}(\Lambda_t \in A|\Lambda_s = \theta), \quad \theta \in \mathcal{S}, A \in \mathcal{B}(\mathcal{S}), \quad (2.12)$$

and

$$Q^\mu(t)h(\theta) = \int_S q(t, \theta, d\gamma)h(\gamma) - q(t, \theta)h(\theta), \quad h \in \mathcal{B}_b(\mathcal{S}), \quad (2.13)$$

where $q(t, \theta, \cdot)$ and $q(t, \theta)$ are given by (2.7), $\mathcal{B}(\mathcal{S})$ denotes the set of measurable functions on $\mathcal{S}$, and $\mathcal{B}_b(\mathcal{S})$ is the set of bounded measurable functions on $\mathcal{S}$.
Proposition 2.4 For a relaxed control \((\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, \Lambda_t, \mu_t, s, \theta)\), it holds, for \(h \in \mathcal{B}_b(S)\),
\[
P_{s,t}^\mu h(\theta) = h(\theta) + \int_s^t Q^\mu(t_1)h(\theta)dt_1 + \int_s^t \int_{t_1}^{t_2} Q^\mu(t_2)Q^\mu(t_1)h(\theta)dt_1dt_2 + \sum_{n=3}^{\infty} \int_s^t \int_{t_n}^{t_2} \cdots \int_s^{t_1} Q^\mu(t_n)Q^\mu(t_{n-1}) \cdots Q^\mu(t_1)h(\theta)dt_1 \cdots dt_{n-1}dt_n.
\]
(2.14)

Proof. Due to (4) of Definition 2.2 and (2.7), (2.8), we know that \((\Lambda_t)\) is a time-inhomogeneous Markov process. Therefore,
\[
P_{s,t+\delta}^\mu h(\theta) = P_{s,t}^\mu P_{t,t+\delta}^\mu h(\theta), \quad h \in \mathcal{B}_b(S).
\]
Invoking (2.8), this yields the equation
\[
\frac{d}{dt} P_{s,t}^\mu h(\theta') = P_{s,t}^\mu Q^\mu(t)h(\theta'), \quad P_{s,s}^\mu h(\theta') = h(\theta'), \quad \theta' \in S, h \in \mathcal{B}_b(S).
\]
(2.15)
See, e.g. [9] for more details on this deduction. Thus, according to [5, Chapter III], formulae (1.12) and (1.15) therein, the unique solution of (2.15) has an explicit representation (2.14) in terms of the Cauchy operator.

Let us show the series in (2.14) is well defined. Endowed with the essential supremum norm \(\| \cdot \|_\infty\), \(\mathcal{B}_b(S)\) becomes a Banach space. Viewed as a linear operator over \(\mathcal{B}_b(S)\), define the operator norm of \(Q^\mu(t)\) by:
\[
\|Q^\mu(t)\| = \sup_{\|h\|_\infty \leq 1} \|Q^\mu(t)h\|_\infty.
\]
which obviously satisfies
\[
\|Q^\mu(t)\| \leq \sup_{\theta \in S} \sup_{u \in U} 2q(\theta; u) \leq 2M < \infty, \quad \forall t \in [0, T].
\]
Hence,
\[
\begin{align*}
\left| \int_s^t \int_{t_n}^{t_2} \cdots \int_s^{t_1} Q^\mu(t_n)Q^\mu(t_{n-1}) \cdots Q^\mu(t_1)h(\theta)dt_1 \cdots dt_{n-1}dt_n \right| \\
\leq \|h\|_\infty \int_s^t \int_{t_n}^{t_2} \cdots \int_s^{t_1} \|Q^\mu(t_n)\| \|Q^\mu(t_{n-1})\| \cdots \|Q^\mu(t_1)\|dt_1 \cdots dt_{n-1}dt_n \\
= \frac{\|h\|_\infty}{n!} \int_s^t \int_{t_n}^{t_2} \cdots \int_s^{t_1} \|Q^\mu(t_n)\| \cdots \|Q^\mu(t_1)\|dt_1 \cdots dt_{n-1}dt_n \\
= \frac{\|h\|_\infty}{n!} \left( \int_s^t \|Q^\mu(r)\|dr \right)^n \\
\leq \frac{(2M(t-s))^n}{n!}\|h\|_\infty,
\end{align*}
\]
(2.16)
since the integral is invariant under any perturbation of the variables \( t_1, \ldots, t_n \). Therefore, the series in (2.14) is convergent, and further the operator \( P_{s,t}^\mu \) is well defined. \( \square \)

Just as done in [17], the relaxed controls can be transformed into controls in the canonical path space to simplify the arguments. Let

\[
\mathcal{U} = \{ \nu : [0, T] \to \mathcal{P}(U); \nu \in \mathcal{D}([0, T]; \mathcal{P}(U)), \ w^\nu_{s,t}(\delta) \leq \psi(\delta), \ \delta \in (0, T) \},
\]

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

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

which is a Polish space endowed with Skorokhod topology. Consider the canonical space

\[
\mathcal{Y} = \mathcal{D}([0, T]; S) \times \mathcal{U}.
\]

Let \( \mathcal{D}, \mathcal{U} \) be their Borel \( \sigma \)-fields, and \( \mathcal{D}_t, \mathcal{U}_t \) the \( \sigma \)-fields up to time \( t \). Put

\[
\tilde{\mathcal{Y}} = \mathcal{D} \times \mathcal{U}, \quad \tilde{\mathcal{Y}}_t = \mathcal{D}_t \times \mathcal{U}_t.
\]

Then, every \( \psi \)-relaxed control \( (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, \Lambda_t, \mu_t, s, \theta) \) can be transformed into a new \( \psi \)-relaxed control \( (\mathcal{Y}, \tilde{\mathcal{Y}}, \tilde{\mathcal{Y}}_t, R, \Lambda_t, \mu_t, s, \theta) \) via the map \( \Psi : \Omega \to \mathcal{Y} \) defined by

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

where \( R = \mathbb{P} \circ \Psi^{-1} \) is a probability measure on \( \mathcal{Y} \). Similar to the discussion in [17], it is clear that the \( \psi \)-relaxed control \( \alpha = (\mathcal{Y}, \tilde{\mathcal{Y}}, \tilde{\mathcal{Y}}_t, R, \Lambda_t, \mu_t, s, \theta) \) is completely determined by the probability measure \( R \), so in the canonical space we use \( R \) itself to denote this \( \psi \)-relaxed control \( \alpha \).

**Proof of Theorem 2.3** If \( V(s, \theta) = \infty \), then every \( \psi \)-relaxed control \( \alpha \) 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 \( \psi \)-relaxed controls on the canonical space, there exists a sequence of probability measures \( R_n, n \geq 1 \), on \( \mathcal{Y} \) such that

\[
\lim_{n \to \infty} J(0, \theta, R_n) = V(0, \theta) < \infty.
\]

In this step, we aim to prove that \( (R_n)_{n \geq 1} \) is tight. To this end, let \( \mathcal{L}_\Lambda^\alpha \) and \( \mathcal{L}_\mu^\alpha, n \geq 1 \), 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, $(\mathcal{P}(U), W_1)$ is a compact Polish space. Then, according to [3, Theorem 14.3] or [8, Theorem 6.3], $U$ is a compact subset in $D([0, T]; \mathcal{P}(U))$. Moreover, by the definition of $\psi$-relaxed control, $\mu$ admits a representation (2.5), and $F_t(\theta')$ is in $U$ for every $\theta' \in S$. The compactness of $\mathcal{P}(U)$ implies the boundedness of $\mathcal{P}(U)$, i.e. there exists a constant $K > 0$ such that $W_1(\nu_1, \nu_2) \leq K$ for any $\nu_1, \nu_2 \in \mathcal{P}(U)$. This yields immediately that for some fixed $\nu$,

$$R_n(\omega : \sup_{0 \leq t \leq T} W_1(\mu_t, \nu) > K) = 0, \quad n \geq 1.$$ 

We go to estimate $R_n(\omega : w''_{\mu(\omega)}(\delta) \geq \varepsilon), n \geq 1$. For any $\varepsilon \in (0, K)$, there exists a $\delta > 0$ such that $\psi(\delta) < \varepsilon$. According to Definition 2.2, for every $\theta' \in S$, denoting by $\nu_t(\cdot, \theta') := \mu_t(\cdot | \Lambda_t = \theta')$, it holds

$$w_{\nu}([t_1, t_2]) \leq \psi(t_2 - t_1) \leq \psi(\delta) < \varepsilon, \quad 0 \leq t_1 < t_2 \leq T, \quad t_2 - t_1 \leq \delta.$$ 

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

$$W_1(\mu_t, \mu_{t_1}) \leq W_1(\nu_t(\cdot, \Lambda_t), \nu_{t_1}(\cdot, \Lambda_t)) + W_1(\nu_{t_1}(\cdot, \Lambda_t), \nu_t(\cdot, \Lambda_t)) \leq W_1(\nu_t(\cdot, \Lambda_t), \nu_{t_1}(\cdot, \Lambda_t)) + K1_{\Lambda_t \neq \Lambda_{t_1}}.$$ 

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

$$\min\{W_1(\mu_{t_1}, \mu_t), W_1(\mu_t, \mu_{t_2})\} \leq \min\{W_1(\nu_t(\cdot, \Lambda_t), \nu_{t_1}(\cdot, \Lambda_t)) + K1_{\Lambda_t \neq \Lambda_{t_1}}, W_1(\nu_{t_1}(\cdot, \Lambda_t), \nu_t(\cdot, \Lambda_t)) + K1_{\Lambda_t \neq \Lambda_{t_2}}\} < \varepsilon.$$ 

Thus,

$$R_n(\omega : \min\{W_1(\mu_{t_1}, \mu_t), W_1(\mu_t, \mu_{t_2})\} \geq \varepsilon) \leq R_n(\omega : \text{the process } (\Lambda_t) \text{ owns at least two jumps during } [t_1, t_2]) \quad (2.19)$$

$$\leq o(\delta).$$

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

$$R_n(\omega : w''_{\mu}(\delta) \geq \varepsilon) \leq o(\delta) \leq \eta. \quad (2.20)$$
For the Markov chain \((\Lambda_t)\) with the bounded transition rate matrices, it is clear that for \(\delta > 0\) sufficiently small,
\[
R_n(\omega : w_\mu([0, \delta)) \geq \varepsilon) \leq \eta, \quad R_n(\omega : w_\mu([T - \delta, T)) \geq \varepsilon) \leq \eta, \quad n \geq 1.
\] (2.21)

Applying [3, Theorem 15.3], we show that \((L_n^\mu)_{n \geq 1}\) is tight.

Next, we go to prove the set of probability measures \((L_n^\Lambda)_{n \geq 1}\) on \(D([0, T]; S)\) is tight. We shall apply Kurtz’s tightness criterion (cf. [8, Theorem 8.6, p.137]) to prove it.

On one hand, by (H4) and Itô’s formula, we have
\[
\mathbb{E}R_n \Phi(\Lambda_t) = \Phi(\theta) + \mathbb{E}R_n \int_0^t Q_{\mu_s} \Phi(\Lambda_s)ds
\leq \Phi(\theta) + \mathbb{E}R_n \int_0^t (\lambda \Phi(\Lambda_s) + \kappa_0)ds.
\]
Then Gronwall’s inequality leads to that
\[
\mathbb{E}R_n \Phi(\Lambda_t) \leq (\Phi(\theta) + \kappa_0 T)e^{\lambda T}, \quad t \in [0, T].
\] (2.22)

Then, for any \(\varepsilon > 0\), take \(N_\varepsilon\) large enough so that
\[
\frac{\mathbb{E}R_n \Phi(\Lambda_t)}{N_\varepsilon} \leq \frac{(\Phi(\theta) + \kappa_0 T)e^{\lambda T}}{N_\varepsilon} < \varepsilon.
\]
Let
\[
K_\varepsilon = \{\gamma \in S; \Phi(\gamma) \leq N_\varepsilon\},
\]
which is a compact set because \(\Phi\) is a compact function. Then,
\[
\sup_n R_n(\Lambda_t \in K_\varepsilon^c) \leq \sup_n \frac{\mathbb{E}R_n \Phi(\Lambda_t)}{N_\varepsilon} < \varepsilon. \quad (2.23)
\]

On the other hand, we also need to show that for any \(\delta > 0\) there exists a nonnegative random variable \(\gamma_n(\delta) \geq 0\) such that
\[
\mathbb{E}R_n \left[1_{\Lambda_{t+u} \neq \Lambda_t} | \mathcal{F}_t \right] \leq \mathbb{E}R_n [\gamma_n(\delta)| \mathcal{F}_t], \quad 0 \leq t \leq T, \quad 0 \leq u \leq \delta,
\]
and \(\lim_{\delta \to 0} \sup_n \mathbb{E}R_n [\gamma_n(\delta)] = 0\). Under (H2), the transition rate \((q(\theta, A; u), q(\theta; u))\) of \((\Lambda_t)\) is bounded, and hence
\[
R_n(\Lambda_s = \Lambda_t, \forall s \in [t, t + u]) \geq \mathbb{E}R_n \left[\exp \left(-\int_t^{t+u} \sup_{\theta \in S} q(\theta; \mu_s)ds \right) \right] \geq \exp(-Mu).
\]
Then, for every $0 \leq u \leq \delta$,
\[
\mathbb{E}_{R_n} \left[ 1_{\{ \Lambda_{t+u} \neq \Lambda_t \}} \right] \leq 1 - R_n(\Lambda_s = \Lambda_t, \forall s \in [t, t+u]) \\
\leq 1 - e^{-Mu} \leq 1 - e^{-M\delta} =: \gamma_n(\delta).
\]
It is clear that $\lim_{\delta \to 0} \sup_n \mathbb{E}_{R_n} \gamma_n(\delta) = 0$. Combining this with (2.23), we conclude that $(\mathcal{L}_n^\alpha)_{n \geq 1}$ is tight.

As a consequence, the fact $(\mathcal{L}_n^\alpha)_{n \geq 1}$ and $(\mathcal{L}_n^\mu)_{n \geq 1}$ are both tight leads to that for any $\varepsilon > 0$, there exist compact sets $K_1 \subset C([0, T]; \mathcal{P}(U))$ and $K_2 \subset \mathcal{D}([0, T]; \mathcal{S})$ such that
\[
R_n(\mathcal{D}([0, T]; \mathcal{S}) \times K_1) = \mathcal{L}_n^\alpha(K_1^c) < \varepsilon,
\]
\[
R_n(K_2 \times \mathcal{P}([0, T] \times U)) = \mathcal{L}_n^\mu(K_2^c) < \varepsilon,
\]
where $K_i^c$, $i = 1, 2$, stands for the complement of $K_i$. So,
\[
R_n((K_1 \times K_2)^c) \leq R_n(\mathcal{D}([0, T]; \mathcal{S}) \times K_1^c) + R_n(K_2^c \times \mathcal{P}([0, T] \times U)) < 2\varepsilon,
\]
which implies the desired tightness of $(R_n)_{n \geq 1}$.

**Step 2.** We go to show the existence of the optimal $\psi$-relaxed control in this step. According to the result of **Step 1**, $(R_n)_{n \geq 1}$ 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. [8, Chapter 3], Theorem 1.8, p. 102), there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ on which are defined $\mathcal{Y}$-valued random variables $Y_n = (\Lambda_t^{(n)}, \mu_t^{(n)})_{t \in [0, T]}$, $n = 1, 2, \ldots$, and $Y_0 = (\Lambda_t^{(0)}, \mu_t^{(0)})_{t \in [0, T]}$ with distribution $R_n, n = 1, 2, \ldots$, and $R_0$ respectively such that
\[
\lim_{n \to \infty} Y_n = Y_0, \quad \mathbb{P}'\text{-a.s.} \quad (2.24)
\]

Denote $\mathcal{F}_t$ the natural $\sigma$-field generated by $(\Lambda_s^{(n)}, \mu_s^{(n)}), n = 0, 1, 2, \ldots$, up to time $t$. We shall prove that $\alpha^* = (\Omega', \mathcal{F}', \mathcal{F}_t, \mathbb{P}', \Lambda_t^{(0)}, \mu_t^{(0)}, 0, \theta)$ is an optimal $\psi$-relaxed control with respect to the value function $V(0, \theta)$. To this end, we need to check that $\alpha^*$ satisfies the conditions of Definition 2.2. Obviously, conditions (1) and (2) of Definition 2.2 hold.

To check condition (4), the transition semigroup of $(\Lambda_t^{(n)}, P_{s,t}^{\mu^{(n)}} 1_A(\theta') := \mathbb{P}'(\Lambda_t^{(n)} \in A|\Lambda_s^{(n)} = \theta'), \theta' \in \mathcal{S}, A \in \mathcal{B}(\mathcal{S})$, is determined by the equation (2.14) with $Q^\mu(t)$ being replaced by $Q^{\mu^{(n)}}(t)$ defined as follows:
\[
Q^{\mu^{(n)}}(t)h(\theta') = \mathbb{E} \left[ \int_{U} \int_{\mathcal{S}} q(\theta', d\gamma; u)h(\gamma)\mu^{(n)}_t(du)|\Lambda_t^{(n)} = \theta' \right] \\
- \mathbb{E} \left[ \int_{U} q(\theta'; u)\mu^{(n)}_t(du)h(\theta')|\Lambda_t^{(n)} = \theta' \right]. 
\quad (2.25)
\]
Similarly, we can define the operators $P_{s,t}^{(0)}$ and $Q_{s,t}^{(0)}(t)$.

For $0 \leq t_1 < t_2 < \ldots < t_k \leq T$, define the projection map $\pi_{t_1 \ldots t_k} : \mathcal{D}([0, T]; S) \to S^k$ by

$$\pi_{t_1 \ldots t_k}(\Lambda) = (\Lambda_{t_1}, \ldots, \Lambda_{t_k}).$$

Let $\mathcal{T}_0$ consist of those $t \in [0, T]$ for which the projection $\pi_t : \mathcal{D}([0, T]; S) \to S$ 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

$$J_t = \{\Lambda \in \mathcal{D}([0, T]; S); \Lambda_t \neq \Lambda_{t-}\}.$$  

Also, $0, T \in \mathcal{T}_0$ by convention. As a probability measure on $\mathcal{D}([0, T]; S)$, it is known that the complement of $\mathcal{T}_0$ in $[0, T]$ is at most countable (cf. [3, p. 124]). Analogously, define the projection map $\tilde{\pi}_{t_1 \ldots t_k} : \mathcal{U} \to \mathcal{P}(U)^k$ by

$$\tilde{\pi}_{t_1 \ldots t_k}(\mu) = (\mu_{t_1}, \ldots, \mu_{t_k}),$$

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 $\mathcal{D}([0, T]; S) \times \mathcal{U}$ as $n \to \infty$ and $\pi_t \times \tilde{\pi}_t$ 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

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

Letting $n \to \infty$ in (2.25) for $t \in \mathcal{T}_0$, we obtain

$$\lim_{n \to \infty} Q_{s,t}^{(n)}(t) h(\theta') = Q_{s,t}^{(0)}(t) h(\theta'), \quad h \in \mathcal{B}_b(S), \quad \theta' \in S.$$  

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

$$\lim_{n \to \infty} \mathbb{P}(\Lambda_t^{(n)} \in A|\Lambda_0^{(n)} = \theta) = \mathbb{P}(\Lambda_t^{(0)} \in A|\Lambda_0^{(0)} = \theta), \quad A \in \mathcal{B}(S), \theta \in S.$$  

Moreover, according to [8, Theorem 7.8, p.131], for every $t \in [0, T]$, there exists a sequence $\{s_n\}_{n \geq 1}$ 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

$$P_{0,s_n}^{(n)} h(\theta') = h(\theta') + \int_0^{s_n} Q_{s_1}^{(n)}(t_1) h(\theta') dt_1 + \int_0^{s_n} \int_0^{t_2} Q_{s_2}^{(n)}(t_2) Q_{s_1}^{(n)}(t_1) h(\theta') dt_1 dt_2$$

$$+ \sum_{k=3}^{\infty} \int_0^{s_n} \int_0^{t_k} \cdots \int_0^{t_2} Q_{s_k}^{(n)}(t_k) Q_{s_{k-1}}^{(n)}(t_{k-1}) \cdots Q_{s_1}^{(n)}(t_1) h(\theta') dt_1 \cdots dt_{k-1} dt_k,$$  

(2.28)
we obtain that
\[
     P_{0,t}^\mu(\theta') = h(\theta') + \int_0^t Q^\mu(t_1)h(\theta')dt_1 + \int_0^t \int_0^{t_2} Q^\mu(t_2)Q^\mu(t_1)h(\theta')dt_1dt_2 \\
     + \sum_{k=3}^{\infty} \int_0^t \int_0^{t_k} \cdots \int_0^{t_{k-1}} Q^\mu(t_k)Q^\mu(t_{k-1}) \cdots Q^\mu(t_1)h(\theta')dt_1 \cdots dt_{k-1}dt_k.
\]

(2.29)

Because the right-hand side of (2.29) is continuous in \( t \), we have from (2.29) that \( t \mapsto P_{0,t}^\mu(\theta') \) is continuous. Whence, (2.8), and equivalently (2.4), is satisfied by taking derivative w.r.t. \( t \) in both sides of (2.29) and taking \( h(\theta') = 1_A(\theta') \) for \( A \in \mathcal{B}(S) \). 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 \( T_0 = [0,T] \).

Now we go to check condition (3). Since \( (\mu_t^{(m)})_{t \in [0,T]} \) converges almost surely to \( (\mu_t^{(0)})_{t \in [0,T]} \) in \( \mathcal{D}([0,T]; \mathcal{P}(U)) \), we have for each \( t \in [0,T] \), \( \mu_t^{(m)} \) converges almost surely to \( \mu_t^{(0)} \) since \( T_0 \) 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 \( \mathcal{F}_{-n}^\Lambda = \sigma(\Lambda_t^{(m)}, m \geq n) \), the completion of the \( \sigma \)-field generated by \( \Lambda_t^{(m)}, m \geq n \). Then

\[
\mathcal{F}_{-1}^\Lambda \supset \mathcal{F}_{-2}^\Lambda \supset \cdots \supset \mathcal{F}_{-n}^\Lambda \supset \mathcal{F}_{-n-1}^\Lambda \supset \cdots.
\]

Put \( \mathcal{F}_{-\infty}^\Lambda = \bigcap_{n \geq 1} \mathcal{F}_{-n}^\Lambda \). \( \mathcal{F}_{-\infty}^\Lambda \) is easily checked to be a \( \sigma \)-field which concerns only the limit behavior of the sequence \( \Lambda_t^{(n)}, n \geq 1 \). Moreover, since there is no point in \( [0,T] \) such that \( (\Lambda_t^{(0)}) \) must jump at that point with positive probability. Therefore, \( \lim_{n \to \infty} \Lambda_t^{(n)} = \Lambda_t^{(0)} \) a.s. for every \( t \in [0,T] \), and further

\[
\mathcal{F}_{-\infty}^\Lambda = \sigma(\Lambda_t^{(0)}).
\]

Define \( \mathcal{F}_n^\mu = \sigma(\mu_t^{(m)}, m \geq n) \). Due to Definition 2.2 (3), \( \mu_t^{(n)} \) is in \( \mathcal{F}_n^\Lambda \) for each \( n \geq 1 \), and hence \( \mathcal{F}_n^\mu \subset \mathcal{F}_n^\Lambda \). Therefore, it follows from the fact \( \lim_{n \to \infty} W_1(\mu_t^{(n)}, \mu_t^{(0)}) = 0 \) a.s. that

\[
\sigma(\mu_t^{(0)}) \subset \bigcap_{n \geq 1} \mathcal{F}_n^\mu \subset \mathcal{F}_{-\infty}^\Lambda = \sigma(\Lambda_t^{(0)}),
\]

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

**Step 3.** Invoking (2.26), (2.24), (2.18), and (2.9), we obtain by the lower semi-
continuity of $f$ and $g$ that

$$V(0, \theta) = \lim_{n \to \infty} \mathbb{E}^{\psi_n} \left[ \int_0^T f(t, \Lambda^{(n)}_t, \mu^{(n)}_t) \, dt + g(\Lambda^{(n)}_T) \right]$$

$$= \lim_{n \to \infty} \mathbb{E}^\psi \left[ \int_0^T \int_U f(t, \Lambda^{(n)}_t, u, \mu^{(n)}_t) \, (du) \, dt + g(\Lambda^{(n)}_T) \right]$$

$$\geq \mathbb{E}^\psi \left[ \int_0^T \int_U f(t, \Lambda^{(0)}_t, u, \mu^{(0)}_t) \, (du) \, dt + g(\Lambda^{(0)}_T) \right]$$

$$\geq V(0, \theta).$$

Hence, $\alpha^*$ is an optimal $\psi$-relaxed control. The proof of this theorem is completed. \qed

After the existence of optimal $\psi$-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; see the argument of Proposition 4.1 in a more complicated situation. 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

$$\varphi'(t, \theta) + f(t, \theta, u) + \sum_{\ell \in \mathcal{S}} q(\theta, \ell; u) \varphi(t, \ell) - q(\theta; u) \varphi(t, \theta) \geq 0,$$

$$\varphi(T, \theta) = g(\theta),$$

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

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

See, for example, [13, 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 \((\Lambda_t)\), which is also called a regime-switching diffusion process; see, [23] and [31] and references therein. The process \((X_t)\) is determined by the following SDE:

\[
dX_t = b(X_t, \Lambda_t)dt + \sigma(X_t, \Lambda_t)dB_t, \tag{3.1}
\]

where \((B_t)\) is a Brownian motion in \(\mathbb{R}^d\); \((\Lambda_t)\) 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

\[
\mathbb{P}(\Lambda_{t+\delta} \in A | \Lambda_t = \theta, \mu_t = \mu) - 1_A(\theta) = (q(\theta, A; \mu) - q(\theta; \mu)1_A(\theta))\delta + o(\delta) \tag{3.2}
\]

provided \(\delta > 0\). The decision-maker still tries to minimize the cost through controlling the transition rates of the Markov chain \((\Lambda_t)\), 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. [16, 17]) or optimal controls for SDEs with regime-switching (see, e.g. [28, 29, 33]), where the control policies are placed directly to the drifts or diffusion coefficients of \((X_t)\). Namely, the controlled system is also given by a SDE

\[
d\tilde{X}_t = b(\tilde{X}_t, \mu_t)dt + \sigma(\tilde{X}_t, \mu_t)dB_t. \tag{3.3}
\]

Roughly speaking, for \((\tilde{X}_t)\), if we change the value of the control \(\mu_t\) at time \(t\), then the speed of \(\tilde{X}_t\) is immediately modified. Nevertheless, for \((X_t)\) given by (3.1), if we change \(\mu_t\) at time \(t\), we only change the switching rate of the process \((\Lambda_t)\) and the speed of \(X_t\) maybe remain the same as before because \(\Lambda_t\) may not jump at \(t\). This observation tells us that in contrast to the process \((\tilde{X}_t)\), the process \((X_t)\) characterized by (3.1) and (3.2) is more closely related to the long time behavior of the control \((\mu_t)\).

Let \(\psi, w^{\mu}_\delta(\delta)\) be defined by (2.2) and (2.3) respectively.

**Definition 3.1** A \(\psi\)-relaxed control is a term \(\alpha = (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B_t, X_t, \Lambda_t, \mu_t, s, x, \theta)\) such that

1. \((s, x, \theta) \in [0, T] \times \mathbb{R}^d \times \mathcal{S}\);
2. \((\Omega, \mathcal{F}, \mathbb{P})\) is a probability space with the filtration \(\{\mathcal{F}_t\}_{t \in [0, T]}\);
3. \((B_t)\) is a \(d\)-dimensional standard Brownian motion on \((\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})\), and \((X_t, \Lambda_t)\) is a stochastic process on \(\mathbb{R}^d \times \mathcal{S}\) satisfying (3.1) and (3.2) with \(X_s = x, \Lambda_s = \theta\);
(4) \( \mu_t \in \mathcal{P}(U) \) is adapted to the \( \sigma \)-field generated by \( \Lambda_t \), \( t \mapsto \mu_t \) is in \( \mathcal{D}([0,T];\mathcal{P}(U)) \) almost surely, and for every \( \theta' \in S \) the curve \( t \mapsto \nu_t(\cdot, \theta') := \mu_t(\cdot | \Lambda_t = \theta') \) satisfies
\[
w_{\nu}([t_1,t_2]) \leq \psi(t_2 - t_1), \quad 0 \leq t_1 < t_2 \leq T;
\]
The collection of all \( \psi \)-relaxed controls with initial value \( (s,x,\theta) \) is denoted by \( \tilde{\Pi}_{s,x,\theta} \).

Given two functions \( f : [0,T] \times \mathbb{R}^d \times S \times U \to \mathbb{R} \) and \( g : \mathbb{R}^d \times S \to \mathbb{R} \), the expected cost relative to the control \( \alpha \in \tilde{\Pi}_{s,x,\theta} \) is defined by
\[
J(s,x,\theta,\alpha) = \mathbb{E}\left[ \int_s^T f(t, X_t, \Lambda_t, \mu_t) dt + g(X_T, \Lambda_T) \right].
\] (3.4)
Correspondingly, the value function is defined by
\[
V(s,x,\theta) = \inf_{\alpha \in \tilde{\Pi}_{s,x,\theta}} J(s,x,\theta,\alpha)
\] (3.5)
for \( s \in [0,T] \), \( x \in \mathbb{R}^d \), \( \theta \in S \). A \( \psi \)-relaxed control \( \alpha^* \in \tilde{\Pi}_{s,x,\theta} \) is called optimal, if it holds
\[
V(s,x,\theta) = J(s,x,\theta,\alpha^*).
\]

We assume that the coefficients of (3.1) satisfy the following conditions.

(H5) There exists a constant \( C_1 > 0 \) such that
\[
|b(x,\theta) - b(y,\theta)|^2 + \|\sigma(x,\theta) - \sigma(y,\theta)\|^2 \leq C_1|x - y|^2, \quad x,y \in \mathbb{R}^d, \theta \in S,
\]
where \( |x|^2 = \sum_{k=1}^d x_k^2 \), \( \|\sigma\|^2 = \text{tr}(\sigma \sigma') \), and \( \sigma' \) is the transpose of the matrix \( \sigma \).

(H6) There exists a constant \( C_2 > 0 \) such that
\[
|b(x,\theta)|^2 + \|\sigma(x,\theta)\|^2 \leq C_2(1 + |x|^2), \quad x \in \mathbb{R}^d, \theta \in S.
\]
The conditions (H5) and (H6) are classical conditions to ensure the existence and uniqueness of nonexplosive solution of SDE (1.1). These conditions can be weakened to include some non-Lipschitz coefficients (cf. e.g. [27]) or singular coefficients (cf. e.g. [20]).

Our second main result of this work is the following theorem.
Theorem 3.2 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 $\psi$-relaxed control $\alpha^* \in \tilde{\Pi}_{s,x,\theta}$.

To simplify the proof, we also transform the relaxed controls into the canonical path space. Let $\mathcal{U}$ be defined by (2.17), and

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

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

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

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

Proof of Theorem 3.2 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 $R_n$, $n \geq 1$, such that

$$\lim_{n \to \infty} J(0, x, \theta, R_n) = V(0, x, \theta) < \infty. \quad (3.7)$$

Step 1. In this step, we aim to prove the tightness of $(R_n)_{n \geq 1}$. Denote by $\mathcal{L}^n_X$, $\mathcal{L}^n_\Lambda$ and $\mathcal{L}^n_{\mu}$, $n \geq 1$, 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 2.3, we can prove the tightness of $(\mathcal{L}^n_\mu)_{n \geq 1}$ and $(\mathcal{L}^n_\Lambda)_{n \geq 1}$. Now, we go to prove the tightness of $(\mathcal{L}^n_X)$. According to [3, Theorem 12.3], it is sufficient to verify the moment condition. By Itô’s formula, for $0 \leq t_1 < t_2 \leq T$,

\[
\begin{align*}
\mathbb{E}_{R_n}[X_{t_2} - X_{t_1}]^4 & \leq 8 \mathbb{E}_{R_n} \left( \int_{t_1}^{t_2} b(X_r, \Lambda_r) dr \right)^4 + 8 \mathbb{E}_{R_n} \left( \int_{t_1}^{t_2} \sigma(X_r, \Lambda_r) dB_r \right)^4 \\
& \leq 8(t_2 - t_1)^3 \mathbb{E}_{R_n} \int_{t_1}^{t_2} |b(X_r, \Lambda_r)|^4 dr + 288(t_2 - t_1)^2 \mathbb{E}_{R_n} \int_{t_1}^{t_2} \|\sigma(X_r, \Lambda_r)\|^4 dr \\
& \leq C(t_2 - t_1) \int_{t_1}^{t_2} (1 + \mathbb{E}_{R_n}|X_r|^4) dr.
\end{align*}
\]
The linear growth condition (H6) implies the existence of a constant $C$ (independent of $n$) such that $\int_0^T \mathbb{E}_{R_n} |X_t|^4 \, dt \leq C$ (cf. [23, Theorem 3.20]). Furthermore, invoking the fact $X_0 = x$, we conclude that $(\mathcal{L}_n^X)_{n \geq 1}$ is tight due to [3, Theorem 12.3].

**Step 2.** Because the marginal distributions of $R_n$, $n \geq 1$ are all tight, we get $R_n$, $n \geq 1$ 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. [8, Chapter 3], Theorem 1.8, p.102) to obtain a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ on which defined a sequence of random variables $(X_t^{(n)}, \Lambda_t^{(n)}, \mu_t^{(n)})_{t \in [0,T]}$; $n \geq 0$, taking values in $\mathcal{Y}$ with the distribution $R_n$, $n \geq 0$, respectively, such that $(X_t^{(n)}, \Lambda_t^{(n)}, \mu_t^{(n)})_{t \in [0,T]}$ converges $\mathbb{P}'$-almost surely to $(X_t^{(0)}, \Lambda_t^{(0)}, \mu_t^{(0)})_{t \in [0,T]}$ as $n \to \infty$.

Let $T_0$ be defined in the same way as the argument of Theorem 2.3. For every $t \in T_0$, we have $(X_t^{(n)}(0), \Lambda_t^{(n)}(0), \mu_t^{(n)}(0))$ converges almost surely to $(X_t^{(0)}(0), \Lambda_t^{(0)}(0), \mu_t^{(0)}(0))$. Analogous to the argument of Theorem 2.3, $(\Lambda_t^{(0)})$ is a continuous time Markov chain with transition rate operator induced from $(\mu_t^{(0)})$, which also implies that $T_0 = [0, T]$. 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 2.3.

We need to check that $(X_t^{(0)})$ satisfies SDE (3.1) under $R_0$ is associated with a $\psi$-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 $\pi_t : C([0,T]; \mathbb{R}^d) \to \mathbb{R}^d$, $\pi_t(X_t) := X_t$, is continuous. Then, this yields that $X_t^{(n)}$ converges almost surely to $X_t^{(0)}$ for each $t \in [0,T]$ 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 $\infty$ in the following integral equation:

$$X_t^{(n)} = x + \int_0^t b(X_s^{(n)}, \Lambda_s^{(n)}) \, ds + \int_0^t \sigma(X_s^{(n)}, \Lambda_s^{(n)}) \, dB_s, \quad (3.9)$$

we get

$$X_t^{(0)} = x + \int_0^t b(X_s^{(0)}, \Lambda_s^{(0)}) \, ds + \int_0^t \sigma(X_s^{(0)}, \Lambda_s^{(0)}) \, dB_s, \quad (3.10)$$

which means that $(X_t^{(0)})$ satisfies SDE (3.1).

Consequently, $R_0$ is a $\psi$-relaxed control. By (3.7) and the lower semi-continuity of $f$ and $g$, we have

$$V(0, x, \theta) = \lim_{n \to \infty} \mathbb{E}_{\mathbb{P}'} \left[ \int_0^T f(t, X_t^{(n)}, \Lambda_t^{(n)}, \mu_t^{(n)}) \, dt + g(X_T^{(n)}, \Lambda_T^{(n)}) \right]$$

21
\[
\begin{align*}
\geq \mathbb{E}_{\bar{\mathcal{F}}'} \left[ \int_0^T f(t, X^0_t, \Lambda^0_t, \mu^0_t) dt + g(X^0_T, \Lambda^0_T) \right] \\
\geq V(0, x, \theta).
\end{align*}
\]

Hence, \(R_0\) is an optimal \(\psi\)-relaxed control. The proof is complete. \(\square\)

## 4 Dynamic programming principle

In this section, we shall establish the dynamic programming principle for the controlled system \((X_t, \Lambda_t)\) introduced in Section 3.

**Proposition 4.1** Suppose the conditions (H1)-(H6) hold. Assume that \(f\) and \(g\) are continuous functions and there exists a positive constant \(C_3\) such that

\[
\begin{align*}
|f(t, x, \theta, u) - f(t, x', \theta, u)| + |g(x, \theta) - g(x', \theta)| &\leq C_3|x - x'|, \\
|f(t, x, \theta, u)| + |g(x, \theta)| &\leq C_3,
\end{align*}
\]

for every \(t \in [0, T]\), \(x, x' \in \mathbb{R}^d\), \(\theta \in \mathcal{S}\) and \(u \in U\). Then the value function \(V(s, x, \theta)\) is a continuous function on \([0, T] \times \mathbb{R}^d \times \mathcal{S}\).

**Proof.** In the following, we use the same notation \(C\) to denote the constants which may change from line to line. Since the topology of \(\mathcal{S}\) is the discrete topology, we only need to consider the continuity of \(V(s, x, \theta)\) in the arguments \(s\) and \(x\).

Let us consider two arbitrary points \((s, x)\) and \((s', x')\) in \([0, T] \times \mathbb{R}^d\). Without loss of generality, assume \(s < s'\). Let \(\theta \in \mathcal{S}\). By Theorem 3.2, there exists an optimal \(\psi\)-relaxed control \(\alpha^* = (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B_t, X_t, \Lambda_t, \mu_t, s, x, \theta)\) such that \(V(s, x, \theta) = J(s, x, \theta, \alpha^*)\).

Let us introduce another \(\psi\)-relaxed control relative to the initial point \((s', x', \theta)\). Define

\[
\Lambda'_t = \Lambda_{t-\Delta s}, \quad \mu'_t = \mu_{t-\Delta s}, \quad \text{where } \Delta s = s' - s.
\]

As a time shift, it is clear that (3.2) holds for \((\Lambda'_t, \mu'_t)\). Consider the following SDE:

\[
dX'_t = b(X'_t, \Lambda'_t) dt + \sigma(X'_t, \Lambda'_t) dB_t, \quad X'_{s'} = x'.
\]

Under the hypothesis (H1), SDE (4.3) admits a unique strong solution \((X'_t)_{t \in [s', T]}\). Then \(\alpha' = (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B_t, X'_t, \Lambda'_t, \mu'_t, s', x', \theta)\) is a \(\psi\)-relaxed control in \(\tilde{\Pi}_{s', x', \theta}\).
By Itô’s formula and (H5),
\[
\mathbb{E}|X'_t - X_t|^2 = \mathbb{E}|x' - X_{s'}|^2 \\
+ \mathbb{E} \int_{s'}^t 2(X'_r - X_r, b(X'_r, \Lambda'_t) - b(X_r, \Lambda_r)) + \|\sigma(X'_r, \Lambda'_t) - \sigma(X_r, \Lambda_r)\|^2 dr \\
\leq \mathbb{E}|x' - X_{s'}|^2 + \mathbb{E} \int_{s'}^t (1 + C_1)|X'_r - X_r|^2 + C(1 + |X'_r|^2 + |X_r|^2)1_{\Lambda'_t \neq \Lambda_r} dr,
\]
where in the last step we have used the linear growth condition (H6), which implies that all the \(p\)-th moments, \(p \geq 1\), of \(X_t\) and \(X'_t\) are finite; cf. e.g. [23]. Applying Gronwall’s inequality, there exists some \(C > 0\) such that
\[
\mathbb{E}|X'_t - X_t|^2 \leq C \left( \mathbb{E}|x' - X_{s'}|^2 + \int_{s'}^T \mathbb{P}(\Lambda'_r \neq \Lambda_r) dr \right) e^{(1+C_1)(t-s')}.
\] (4.4)

By virtue of (3.2) and (H2), we have that for any \(r \in [0, T]\),
\[
\mathbb{P}(\Lambda'_r \neq \Lambda_r) = \mathbb{P}(\Lambda_r - \Delta_s \neq \Lambda_r) \leq M \Delta s + o(\Delta s).
\] (4.5)

Direct calculation leads to
\[
\mathbb{E}|x' - X_{s'}|^2 \leq 2\mathbb{E}|x - X_{s'}|^2 + 2|x' - x|^2 \\
\leq 2|x' - x|^2 + 4(s' - s) \int_s^{s'} \mathbb{E}|b(X_r, \Lambda_r)|^2 dr + 4\mathbb{E} \int_s^{s'} \|\sigma(X_r, \Lambda_r)\|^2 dr,
\]
which converges to 0 as \(s' \downarrow s\). Inserting this and (4.5) into (4.4), we have
\[
\lim_{(s',x') \to (s,x)} \mathbb{E}|X'_t - X_t|^2 = 0.
\] (4.6)

By the definition of value function, we have
\[
V(s', x', \theta) - V(s, x, \theta) \\
\leq \mathbb{E} \left[ \int_{s'}^T (f(t, X'_t, \Lambda'_t, \mu'_t) - f(t, X_t, \Lambda_t, \mu_t)) dt + g(X'_T, \Lambda'_T) - g(X_T, \Lambda_T) \right] \\
\leq \mathbb{E} \left[ \int_{s'}^T (f(t, X'_t, \Lambda'_t, \mu'_t) - f(t, X_t, \Lambda'_t, \mu'_t)) + (f(t, X_t, \Lambda'_t, \mu'_t) - f(t, X_t, \Lambda_t, \mu_t)) dt \\
+ (g(X'_T, \Lambda'_T) - g(X_T, \Lambda'_T)) + (g(X_T, \Lambda'_T) - g(X_T, \Lambda_T)) \right] \\
\leq \mathbb{E} \left[ \int_{s'}^T C_3|X'_t - X_t| dt + 2C_31_{\Lambda'_T \neq \Lambda_T} \right] + \mathbb{E} \left[ \int_{s'}^T f(t, X_t, \Lambda'_t, \mu'_t) - f(t, X_t, \Lambda_t, \mu_t) dt \right].
\]
To estimate the last term, by the definition of $\Lambda^t$ and $\mu^t$, it holds
\[
E \left[ \int_{s'}^T (f(t, X_t, \Lambda^t, \mu^t) - f(t, X_t, \Lambda, \mu)) dt \right]
= E \left[ \int_s^{s+\Delta s} f(t, X_t+\Delta s, \Lambda^t, \mu^t) dt - \int_s^T f(t, X_t, \Lambda, \mu) dt \right]
\]
\[
= E \left[ \int_s^{s'} f(t, X_t, \Lambda, \mu) dt - \int_{s+\Delta s}^T f(t, X_t, \Lambda, \mu) dt \right]
+ E \left[ \int_s^{T-\Delta s} (f(t+\Delta s, X_{t+\Delta s}, \Lambda, \mu) - f(t, X_t, \Lambda, \mu)) dt \right].
\]

Therefore, by the dominated convergence theorem and the boundedness of $f$, we obtain
\[
\lim_{s' \to s} E \left[ \int_{s'}^T (f(t, X_t, \Lambda^t, \mu^t) - f(t, X_t, \Lambda, \mu)) dt \right] = 0.
\]

Combining this with (4.5), (4.6) and (4.7), we have
\[
\lim_{(s', x') \to (s, x)} V(s', x', \theta) - V(s, x, \theta) \leq 0.
\]

Then, using the symmetric position of $(s', x')$ and $(s, x)$, we can get
\[
\lim_{(s', x') \to (s, x)} V(s', x', \theta) - V(s, x, \theta) = 0,
\]
and conclude the proof. \[\square\]

According to Theorem 3.2, for every $(s, x, \theta) \in [0, T] \times \mathbb{R}^d \times \mathcal{S}$, there is an optimal $\psi$-relaxed control $R$ on the canonical space $\mathcal{Y}$ defined by (3.6). Next, we shall prove that there exists a measurable map $(s, x, \theta) \mapsto H(s, x, \theta)$ from $[0, T] \times \mathbb{R}^d \times \mathcal{S}$ to $\tilde{\Pi}_{s, x, \theta}$ such that $H(s, x, \theta)$ is an optimal $\psi$-relaxed control in $\tilde{\Pi}_{s, x, \theta}$. To this end, similar to [17], we adopt the idea and notation of Stroock and Varadhan [30] in the study of measurable choices on separable metric space.

As a product space, on $\mathcal{Y}$, we introduce the following metric which makes $\mathcal{Y}$ to be a Polish space:
\[
\rho((x, \Lambda, \mu), (x', \Lambda', \mu')) = \|x - x'\|_{\infty} + d(\Lambda, \Lambda') + d(\mu, \mu')
\] (4.8)
for $(x, \Lambda, \mu), (x', \Lambda', \mu')$ in $\mathcal{Y}$, where
\[
\|x - x'\|_{\infty} = \max_{t \in [0, T]} |x_t - x'_t|.
\]
and $d$ is the metric defined in [8, p.117] on the càdlàg space $D([0,T];E)$ over a general Polish space $E$, which makes $D([0,T];E)$ to be a Polish space. Let

$$\mathcal{R}_{s,x,\theta}^0 = \{ R \in \tilde{\Pi}_{s,x,\theta}; J(s,x,\theta,R) = V(s,x,\theta) \}, \quad (s,x,\theta) \in [0,T] \times \mathbb{R}^d \times \mathcal{S}. \quad (4.9)$$

It is known (cf. [1]) that $\mathcal{P}(Y)$ is a Polish space endowed with $L^1$-Wasserstein distance $W_{1,Y}$, which is defined as follows: for any $\tilde{R}_1$ and $\tilde{R}_2$ in $\mathcal{P}(Y)$,

$$W_{1,Y}(\tilde{R}_1, \tilde{R}_2) = \inf_{\Gamma \in \mathcal{C}(\tilde{R}_1,\tilde{R}_2)} \left\{ \int_{Y \times Y} \rho((x,\Lambda,\mu),(x',\Lambda',\mu'))d\Gamma \right\}.$$

As a subset of $\mathcal{P}(Y)$, $\mathcal{R}_{s,x,\theta}^0$ is closed under the metric $W_{1,Y}$. Analogous to the argument of Theorem 3.2, we can show that $\mathcal{R}_{s,x,\theta}^0$ is tight. By Prohorov’s theorem, $\mathcal{R}_{s,x,\theta}^0$ is a compact set in $\mathcal{P}(Y)$.

Denote by $\text{Comp}(\mathcal{P}(Y))$ the space of all compact subsets of $\mathcal{P}(Y)$, and define a metric $\text{dist}(K_1,K_2)$ between two points $K_1$, $K_2$ of $\text{Comp}(\mathcal{P}(Y))$ by

$$\text{dist}(K_1,K_2) = \inf \{ \varepsilon > 0; K_1 \subset K_1^\varepsilon, K_2 \subset K_1^\varepsilon \}.$$

Here, for all set $A \in \text{Comp}(\mathcal{P}(Y))$,

$$A^\varepsilon := \{ \tilde{\nu} \in \mathcal{P}(Y); W_{1,Y}(\tilde{\nu},\tilde{\mu}) < \varepsilon \text{ for some } \tilde{\mu} \in A \}.$$

**Proposition 4.2** Assume all the assumptions of Proposition 4.1 hold. Then $\mathcal{R}_0 : [0,T] \times \mathbb{R}^d \times \mathcal{S} \to \text{Comp}(\mathcal{P}(Y))$ is Borel measurable. Moreover, there exists a measurable selector $H$ of $\mathcal{R}_0$, i.e. $H(s,x,\theta) \in \mathcal{R}_{s,x,\theta}^0$, and $H : [0,T] \times \mathbb{R}^d \times \mathcal{S} \to \mathcal{P}(Y)$ is Borel measurable.

**Proof.** According to [30, Lemma 12.1.8], we need only to show that for $(s_n,x_n,\theta_n) \in [0,T] \times \mathbb{R}^d \times \mathcal{S}$ with $(s_n,x_n,\theta_n) \to (s,x,\theta)$ as $n \to \infty$, there exists a subsequence $R_{n_k} \in \mathcal{R}_{s_{n_k},x_{n_k},\theta_{n_k}}^0$ and $R_0$ in $\mathcal{R}_{s,x,\theta}^0$ such that $R_{n_k}$ converges weakly to $R_0$ as $k \to \infty$.

In fact, it is clear that the proof for the tightness on $\mathcal{L}_X^n$ and $\mathcal{L}_\mu^n$ in Theorem 3.2 is still valid for $R_n$. Some modification is needed to prove the tightness of $\mathcal{L}_X^n$. Indeed, at current situation, the moment condition (3.8) still holds. Now, noting that the initial value $x_n$ converges to $x$, applying [3, Theorem 12.3], $(\mathcal{L}_X^n)_{n \geq 1}$ is tight. Therefore, this leads to the tightness of $(R_n)_{n \geq 1}$. Furthermore, there exists a subsequence $R_{n_k}$ and $R_0 \in \mathcal{P}(Y)$ such that $R_{n_k}$ converges weakly to $R_0$ as $k \to \infty$. We also need to show $R_0 \in \mathcal{R}_{s,x,\theta}^0$. 

25
Since \((s_n, x_n, \theta_n) \to (s, x, \theta)\) as \(n \to \infty\), it is clear that \(R_0 \in \tilde{\Pi}_{s,x,\theta}\). Meanwhile, due to Proposition 4.1 and the lower semi-continuity of \(f\) and \(g\), we have

\[
V(s, x, \theta) = \lim_{k \to \infty} V(s_{n_k}, x_{n_k}, \theta_{n_k}) \\
= \mathbb{E}_{R_{n_k}} \left[ \int_s^T f(t, X_t, \Lambda_t, \mu_t)dt + g(X_T, \Lambda_T) \right] \\
\geq \mathbb{E}_{R_0} \left[ \int_s^T f(t, X_t, \Lambda_t, \mu_t)dt + g(X_T, \Lambda_T) \right] \\
\geq V(s, x, \theta).
\]

This means that \(R_0 \in \mathcal{R}^0_{s,x,\theta}\).

The existence of a measurable selector \(H\) is a consequence of a selection theorem [30, Theorem 12.1.10].

With the preparation of Proposition 4.2, we can now establish the dynamic programming principle. We adopt the notation of [18]. For an \(\mathcal{F}_t\)-stopping time \(\tau\), \(\mathcal{F}_\tau\) is the collection of sets \(A\) such that \(A \cap \{\tau \leq t\} \in \mathcal{F}_t\). According to [30, Theorem 6.1.2] or [18, Lemma 3.3], for an \(\mathcal{F}_t\)-stopping time \(\tau\) with \(s \leq \tau \leq T\), and a probability measure \(\mathbb{P}\) on \(\mathcal{Y}\) belonging to \(\tilde{\Pi}_{s,x,\theta}\), there exists a probability measure \(\mathbb{P} \otimes_\tau H\) on \(\mathcal{Y}\) such that

\[
\forall A \in \mathcal{F}_\tau, \quad \mathbb{P} \otimes_\tau H (A) = \mathbb{P}(A),
\]

and \(\mathbb{P} \otimes_\tau H \in \tilde{\Pi}_{s,x,\theta}\) by combining [18, Proposition 3.8] with Proposition 4.2.

**Theorem 4.3** Assume all the conditions of Proposition 4.1 are still valid. If \(\tau\) is an \(\mathcal{F}_t\)-stopping time with \(s \leq \tau \leq T\), then

\[
V(s, x, \theta) = \inf \left\{ \mathbb{E}_\mathbb{P} \left[ \int_s^T f(t, X_t, \Lambda_t, \mu_t)dt + V(\tau, X_\tau, \Lambda_\tau) \right]; \ \mathbb{P} \in \tilde{\Pi}_{s,x,\theta} \right\}. \tag{4.10}
\]

**Proof.** Let \(\mathbb{P} \in \tilde{\Pi}_{s,x,\theta}\), then \(\mathbb{P} \otimes_\tau H \in \tilde{\Pi}_{s,x,\theta}\). Due to the definition of \(V\),

\[
V(s, x, \theta) \\
\leq \mathbb{E}_{\mathbb{P} \otimes_\tau H} \left[ \int_s^T f(t, X_t, \Lambda_t, \mu_t)dt + g(X_T, \Lambda_T) \right] \\
= \mathbb{E}_{\mathbb{P} \otimes_\tau H} \left[ \int_s^\tau f(t, X_t, \Lambda_t, \mu_t)dt + \int_\tau^T f(t, X_t, \Lambda_t, \mu_t)dt + g(X_T, \Lambda_T) \right]
\]
\[= \mathbb{E}_{\mathbb{P} \otimes \mathcal{H}} \left[ \int_{s}^{T} f(t, X_t, \Lambda_t, \mu_t) \, dt + \mathbb{E}_{\mathbb{P} \otimes \mathcal{H}} \left[ \int_{\tau}^{T} f(t, X_t, \Lambda_t, \mu_t) \, dt + g(X_T, \Lambda_T) \bigg| \mathcal{F}_\tau \right] \right] \]

\[= \mathbb{E}_{\mathbb{P}} \left[ \int_{s}^{\tau} f(t, X_t, \Lambda_t, \mu_t) \, dt + V(\tau, X_\tau, \Lambda_\tau) \right]. \]

The arbitrariness of \(P \in \tilde{\Pi}_{s,x,\theta}\) yields that

\[V(s, x, \theta) \leq \inf \left\{ \mathbb{E}_P \left[ \int_{s}^{\tau} f(t, X_t, \Lambda_t, \mu_t) \, dt + V(\tau, X_\tau, \Lambda_\tau) \right] ; \ P \in \tilde{\Pi}_{s,x,\theta} \right\}. \]

On the other hand, by Theorem 3.2, there exists an optimal \(\psi\)-relaxed control \(\bar{P}\) on \(\mathcal{Y}\). Then, we have

\[V(s, x, \theta) = \mathbb{E}_{\bar{P}} \left[ \int_{s}^{T} f(t, X_t, \Lambda_t, \mu_t) \, dt + g(X_T, \Lambda_T) \right] \]

\[= \mathbb{E}_{\bar{P}} \left[ \int_{s}^{\tau} f(t, X_t, \Lambda_t, \mu_t) \, dt + \int_{\tau}^{T} f(t, X_t, \Lambda_t, \mu_t) \, dt + g(X_T, \Lambda_T) \right] \]

\[\geq \mathbb{E}_{\bar{P}} \left[ \int_{s}^{\tau} f(t, X_t, \Lambda_t, \mu_t) \, dt + V(\tau, X_\tau, \Lambda_\tau) \right] \]

\[\geq \inf \left\{ \mathbb{E}_P \left[ \int_{s}^{\tau} f(t, X_t, \Lambda_t, \mu_t) \, dt + V(\tau, X_\tau, \Lambda_\tau) \right] ; \ P \in \tilde{\Pi}_{s,x,\theta} \right\} \]

Consequently, we establish the dynamical programming principle (4.10). \(\square\)

In the end of this work, we provide an heuristic derivation of the dynamic programming equation under the assumption that the value function \(V(s, x, \theta)\) satisfies the regularity condition of Itô’s formula. Denote the infinitesimal generator of the process \((X_t, \Lambda_t)\) by \(\mathcal{A}\) defined by:

\[\mathcal{A}h(t, x, \theta) = \frac{\partial h}{\partial t} + \sum_{i,j=1}^{d} a_{ij}(x, \theta) \frac{\partial^2 h}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i(x, \theta) \frac{\partial h}{\partial x_i} + Q^\mu(t)h(t, x, \theta),\]

where \((a_{ij})(x, \theta) = \sigma(x, \theta)\sigma(x, \theta)^*\) and the operator \(Q^\mu(t)\) is defined by (2.13). Applying Itô’s formula to \(V(s, x, \theta)\), we obtain

\[\mathbb{E}_P V(t, X_t, \Lambda_t) = V(s, x, \theta) + \mathbb{E}_P \left[ \int_{s}^{t} \mathcal{A}V(r, X_r, \Lambda_r) \, dr \right] \]

for \(P \in \tilde{\Pi}_{s,x,\theta}\). Inserting this equality into (4.10), we get

\[\inf_{P \in \tilde{\Pi}_{s,x,\theta}} \mathbb{E}_P \left[ \int_{s}^{t} (\mathcal{A}V(r, X_r, \Lambda_r) + f(r, X_r, \Lambda_r)) \, dr \right] = 0. \quad (4.11)\]
As is well known in classical control problem, the value function is a solution of the corresponding Hamilton-Jacobi-Bellman equation when it has sufficient regularity, and usually it is only a viscosity solution.

References

[1] L. Ambrosio, N. Gigli, G. Savaré, Gradient flows in metric spaces and in the space of probability measures, Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.

[2] J. Bao, J. Shao, Permance and extinction of regime-switching predator-prey models, SIAM J. Math. Anal. 48 (2016), 725-739.

[3] P. Billingsley, convergence of probability measures, John Wiley, New York, 1968.

[4] P.L. Chow, J. Menaldi, M. Robin, Additive control of stochastic linear system with finite time horizon, SIAM J. Control Optim. 23 (1985), 858-899.

[5] J. Daleckij, M. Krein, Stability of solutions of differential equations in Banach space, Amer. Math. Soc. Transl. 43, 1974.

[6] N.H. Du, H.N. Dang, G. Yin, Conditions for permanence and ergodicity of certain stochastic predator-prey models, J. Appl. Probab. 53 (2015), 187-202.

[7] F. Dufour, B. Miller, Maximum principle for stochastic control problems, SIAM J. Control Optim. 45 (2006), 668-698.

[8] S. Ethier, T. Kurtz, Markov processes characterization and convergence, Wiley, New York, 1986.

[9] W. Feller, On the integro-differential equations of purely discontinuous Markoff processes. Trans. Amer. Math. Soc. 48 (1940), 488-515.

[10] X.P. Guo, Continuous-time Makrov decision processes with discounted rewards: the case of Polish spaces, Math. Oper. Res. 32 (2007), 73-87.

[11] X.P. Guo, O. Hernández-Lerma, Continuous-time controlled Markov chains, Ann. Appl. Prob. 13 (2003), 363-388.
[12] X.P. Guo, O. Hernández-Lerma, Continuous-time Markov decision processes. Theory and applications, Springer-Verlag, Berlin, 2009.

[13] X.P. Guo, X. Huang, Y. Huang, Finite-horizon optimality for continuous-time Markov decision processes with unbounded transition rates, Adv. Appl. Prob. 47 (2015), 1064-1087.

[14] X.P. Guo, U. Rieder, Average optimality for continuous-time Markov decision processes in Polish spaces, Ann. Appl. Probab. 16 (2006), 730-756.

[15] X.P. Guo, M. Vykertas, Y. Zhang, Absorbing continuous-time Markov decision processes with total cost criteria, Adv. Appl. Prob. 45 (2013), 490-519.

[16] U. Haussmann, J. Lepeltier, On the existence of optimal control, SIAM J. Control Optim. 28 (1990), 851-902.

[17] U. Haussmann, W. Suo, Singular optimal stochastic controls I: existence, SIAM J. Control Optim. 33 (1995), 916-936.

[18] U. Haussmann, W. Suo, Singular optimal stochastic controls II: Dynamic programming, SIAM J. Control Optim. 33 (1995), 937-959.

[19] I. Karatzas, S. Shreve, Connections between optimal stopping and stochastic control I: Monotone follower problem, SIAM J. Control Optim. 22 (1984), 856-877.

[20] N.V. Krylov, M. Röckner, Strong solutions of stochastic equations with singular time dependent drift, Probab. Theory Related Fields 131 (2005), 154-196.

[21] H.J. Kushner, Existence results for optimal stochastic controls, J. Optim. Theory Appl., 15 (1975), 347-359.

[22] X. Mao, Stabilization of continuous-time hybrid stochastic differential equations by discrete time feedback control, Automatica J. IFAC, 49 (2013), 3677-3681.

[23] X. Mao, C. Yuan, Stochastic Differential Equations with Markovian Switching, Imperial College Press, London, 2006.

[24] B. Miller, Finite state continuous time Markov decision processes with a finite planning horizon, J. Math. Anal. Appl. 22 (1968), 552-569.
[25] A. Piunovskiy, Y. Zhang, Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach, SIAM J. Control Optim. 49 (2011), 2032-2061.

[26] T. Prieto-Rumeau, O. Hernández-Lerma, Selected topics on continuous-time controlled Markov chains and Markov games, Imperial College Press, London, 2012.

[27] J. Shao, Strong solutions and strong Feller properties for regime-switching diffusion processes in an infinite state space, SIAM J. Control Optim. 53 (2015), 2462-2479.

[28] Q. Song, R. Stockbridge, C. Zhu, On optimal harvesting problems in random environments, SIAM J. Control Optim. 49 (2011), 859-889.

[29] Q. Song, C. Zhu, On singular control problems with state constraints and regime-switching: a viscosity solution approach, Automatica J. IFAC 70 (2016), 66-73.

[30] D.W. Stroock, S.R.S. Varadhan, Multidimensional diffusion processes, Springer-Verlag, New York, 1979.

[31] G. Yin, C. Zhu, Hybrid switching diffusions: properties and applications, Vol. 63, Stochastic Modeling and Applied Probability, Springer, New York. 2010.

[32] A. Yushkevich, Controlled Markov models with countable state space and continuous time, Theory Prob. Appl. 22 (1978), 215-235.

[33] X. Zhou, G. Yin, Markowitz’s mean-variance portfolio section with regime-switching: A continuous time model, SIAM J. Control Optim. 42 (2003), 1466-1482.