Existence of optimal delay-dependent control for finite-horizon continuous-time Markov decision process

Zhong-Wei Liao † Jinghai Shao‡

April 1, 2020

Abstract

This paper intends to study the optimal control problem for the continuous-time Markov decision process with denumerable states and compact action space. The admissible controls depend not only on the current state of the jumping process but also on its history. By the compactification method, we show the existence of an optimal delay-dependent control under some explicit conditions, and further establish the dynamic programming principle. Moreover, we show that the value function is the unique viscosity solution of certain Hamilton-Jacobi-Bellman equation which does not depend on the delay-dependent control policies. Consequently, under our explicit conditions, there is no impact on the value function to make decision depending on or not on the history of the jumping process.

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

Keywords: Markov decision process, finite-horizon criterion, delay-dependent control, compactification method

*Supported in part by NNSFs of China (No. 11771327, 11831014, 11701588)
†South China Research Center for Applied Mathematics and Interdisciplinary Studies, South China Normal University, China. Email: zhwliao@hotmail.com
‡Center for Applied Mathematics, Tianjin University, Tianjin 300072, China. Email: shaojh@tju.edu.cn.
1 Introduction

In this paper, we show the existence of an optimal policy out of the class of randomized delay-dependent controls for finite-horizon continuous-time Markov decision processes (CTMDPs) with denumerable state space and compact action space. CTMDPs have been studied intensively due to their rich application in queueing systems, population processes, see, e.g. the monographs [2, 9, 22], the recent works [8, 10, 11, 14, 20, 23], and the extensive references therein. For the actual controlled model, the phenomenon of delay is everywhere and cannot be ignored. It means that the decision maker cannot make decisions based on the current state of the system, but on the state before a strictly positive time. To our best knowledge, there is a lack of work has been done on these delay phenomenons and we aim at presenting an appropriate mathematical model to describe the control problems with delay.

As is well know, the expected finite-horizon criterion has been a common optimality criterion for CTMDPs optimization problems, which has been studied by numerous authors, see e.g. [2, 8, 10, 19, 21, 29]. For finite-horizon CTMDPs with finite state and action space, Miller [19] gave a necessary and sufficient condition for the existence of a piecewise constant optimal policy. Subsequently, the state space of CTMDPs had been generalized to denumerable space (cf. [29]) and Borel space (cf. [21]), and the existence of an optimal Markov policy had been proven under the bounded hypothesis of transition and cost rates. Recently, Baüerle and Rieder [2] studies the finite-horizon CTMDPs with Markov polices by the method that based on the equivalent transformation from finite-horizon CTMDPs to infinite-horizon discrete-time Markov decision processes. The corresponding optimality equation had been established according to the existing theory on discrete-time Markov decision processes. In addition, Ghosh and Saha [8] considered the finite-horizon CTMDPs in Borel state space with bounded transition rates and Markov policies. The existence of a unique solution of the optimality equation is guaranteed by the Banach fixed point theorem, relatively, the existence of an optimal Markov policy is based on the Itô-Dynkin’s formula. The finite-horizon CTMDPs with unbounded transition and cost rates are investigated in Guo et al. [10], in which the existence of optimal Markov policies has been proven.

The present paper deals with the finite-horizon CTMDPs with denumerable state space and compact metric action space, but is rather different from the aforementioned works [2, 8, 10]. The main contributions of the present paper are as follows:

(i) In comparison with [8, 10], our method used in the existence of an optimal delay-dependent control does not involve the solvability of the differential equation, but is based on
the compactification method, which is an effective method in the research of optimal control problem of jump-diffusion processes, cf. [4, 5, 12, 13]. The basic idea is inspired by Kushner [16], Haussmann and Suo [12, 13], and recently Shao and Zhao [25], Shao [26]. [25] only studied the Markov control policies, we can deal with delay-dependent control policies. Due to the appearance of delays in the control, the studied system is not a Markovian process any more. Our approach is also suitable to the other case of the optimality criteria in CTMDPs, such as expected discounted, average and risk-sensitive.

(ii) According to the measurable selection theorem (cf. e.g. Stroock and Varadhan [27]), the dynamic programming principle of CTMDPs is established in Theorem 4.1. For the classical Markov decision processes, the dynamic programming principle can induce that the value function is a solution to the corresponding Hamilton-Jacobi-Bellman (HJB) equation when it admits appropriate regularity; see e.g. [8, 9, 10, 15]. Usually, it is hard to check the desired regularity of the value function, and the value function is only a viscosity solution of the HJB equation (see [7] and the references therein). In current situation, under some explicit conditions we show that the value function is Lipschitz continuous, and is the unique viscosity solution to the corresponding HJB equation.

(iii) The derived HJB equation does not depend on whether the control policies are delay-dependent or not. So, there is no improvement on the value function to make decisions depending on the history of the jumping process under our explicit conditions.

The rest of our paper is organized as follows. In Section 2, we state the concept of delay-dependent control and the optimality problem that we are concerned with. For the convenience, the optimality problem is reformulated on the canonical path space. In Section 3, we prove the existence of the optimal delay-dependent control of our model. In Section 4, we prove the continuity of the value function and establish the dynamic programming principle. Then we show that the value function is a unique viscosity solution of certain HJB equation.

2 Formulation and assumptions

The objective of this section is to describe briefly the controlled process and the associated optimal control criterion of interest in this paper. Let \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) be a filtered probability space satisfying the usual conditions, i.e. \((\Omega, \mathcal{F}, \mathbb{P})\) is complete, the filtration \((\mathcal{F}_t)_{t \geq 0}\) is right-continuous and \(\mathcal{F}_0\) contains all \(\mathbb{P}\)-null sets in \(\mathcal{F}\). Let \(S = \{1, 2, \ldots\}\) be the
countable state space, $U$ be the action space which is a compact subset of $\mathbb{R}^k$ for some $k \in \mathbb{N}$. Denote by $\mathcal{P}(U)$ the collection of all probability measures over $U$, which is endowed with $L_1$-Wasserstein distance $W_1$ defined by:

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

where $\mathcal{C}(\mu, \nu)$ stands for the set of all couplings of $\mu$ and $\nu$ in $\mathcal{P}(U)$. Since $U$ is compact, $\mathcal{P}(U)$ becomes a compact Polish space under the metric $W_1$, and the weak convergence of probability measures in $\mathcal{P}(U)$ is equivalent to the convergence in the $W_1$ distance (cf. e.g. [1, Chapter 7]). In this work we investigate finite-horizon optimal control problem on $[0, T]$, where $T > 0$ is fixed throughout this work.

For each $\mu \in \mathcal{P}(U)$, $(q_{ij}(\mu))$ is a transition rate matrix over the state space $S$, which is assumed to be conservative, i.e.

$$\sum_{j \neq i} q_{ij}(\mu) = q_i(\mu) = -q_{ii}(\mu), \; \forall i \in S, \; \mu \in \mathcal{P}(U).$$

The process $(\Lambda_t)$ is an $\mathcal{F}_t$-adapted jump process on $S$ satisfying

$$\mathbb{P}(\Lambda_{t+\delta} = j | \Lambda_t = i, \mu_t = \mu) = \begin{cases} q_{ij}(\mu)\delta + o(\delta), & \text{if } i \neq j, \\ 1 + q_{ii}(\mu)\delta + o(\delta), & \text{otherwise,} \end{cases} \quad (2.1)$$

provided $\delta > 0$.

In order to introduce the delay-dependent control, we first introduce some notations. Given any metric space $E$, denote by $\mathcal{C}([0, T]; E)$ the collection of continuous functions $x : [0, T] \to E$, and $\mathcal{D}([0, T]; E)$ the collection of right-continuous functions with left limits $\lambda : [0, T] \to E$. For $r_0 \in (0, T)$ and $s \in [0, T]$, define a shift operator $\theta_{s, r_0} : \mathcal{D}([0, T]; S) \to \mathcal{D}([0, T]; S)$ by

$$(\theta_{s, r_0} \lambda)(t) = \lambda((t - r_0) \vee s), \; t \in [0, T].$$

Moreover, $\theta_{s, r_0}^k \lambda(t) := \lambda((t - kr_0) \vee s)$ for $\lambda \in \mathcal{D}([0, T]; S), \; k \in \mathbb{Z}_+$. Next, we introduce the concept of delay-dependent control.

**Definition 2.1** Fixed any $m \in \mathbb{Z}_+$ and $r_0 > 0$. Given any $s \in [0, T)$ and $i \in S$. A randomized delay-dependent control is a term $\alpha = (\Lambda_t, \mu_t, s, i)$ such that

(i) $(\Lambda_t)$ is an $\mathcal{F}_t$-adapted jump process satisfying (2.1) with initial value $\Lambda_s = i$.  

---

4
(ii) There exists a measurable map $h : [0, T] \times S^{m+1} \to \mathcal{P}(U)$ such that

$$\mu_t = h(t, \theta_{s,r_0}^0 \Lambda(t), \ldots, \theta_{s,r_0}^m \Lambda(t)), \quad \text{almost all } t \in [s, T]. \quad (2.3)$$

The parameter $r_0 > 0$ is used to characterize the time interval of delay of the controlled processes, and $m \in \mathbb{Z}_+$ for the number of delay. The collection of all delay-dependent control $\alpha$ with initial condition $(s, i)$ is denoted by $\Pi_{s,i}$. When the starting time of the optimal control problem is $s$, as we have no further information on the controlled system before initial time $s$, we use the state of the process $(\Lambda_t)$ at time $s$ to represent its states before time $s$, which is reflected by the definition of $\mu_t$ through equation (2.3). Such treatment has been used in the study of optimal control problem over history-dependent policies; see, for instance, [10, 11].

Let $f : [0, T] \times S \times \mathcal{P}(U) \to [0, \infty)$, $g : S \to [0, \infty)$ be two lower semi-continuous functions. The expected cost for the delay-dependent control $\alpha \in \Pi_{s,i}$ is defined by

$$J(s, i, \alpha) = \mathbb{E} \left[ \int_s^T f(t, \Lambda_t, \mu_t)dt + g(\Lambda_T) \right], \quad (2.4)$$

and the value function is defined by

$$V(s, i) = \inf_{\alpha \in \Pi_{s,i}} J(s, i, \alpha). \quad (2.5)$$

It immediately implies that the value function $V$ satisfies $V(T, i) = g(i), \forall i \in S$. A delay-dependent control $\alpha^* \in \Pi_{s,i}$ is said to be optimal, if $V(s, i) = J(s, i, \alpha^*)$.

The set of delay-dependent controls introduced in Definition 2.1 contains many interesting control policies. Next, we present some examples below.

**Example 2.1** We consider the optimal control problem with initial time $s = 0$.

1. $\mu_t = h(\Lambda_t)$ for some $h : S \to \mathcal{P}(U)$. In this situation, $\alpha$ is corresponding to the stationary randomized Markov policy studied by many works; see, e.g. [9].

2. $\mu_t = h(\Lambda_{(t-r_0)\vee 0}, \Lambda_{(t-2r_0)\vee 0})$ for some measurable $h : S \to \mathcal{P}(U)$. Now the control policies are purely determined by the jump process with a positive delay. This kind of controls is very natural to be used in the realistic application.

3. $\mu_t = h(t, \Lambda_{(t-r_0)\vee 0}, \Lambda_{(t-2r_0)\vee 0})$ for some measurable $h : [0, T] \times S \times S \to \mathcal{P}(U)$. 

4. \( \mu_t = h(t, \Lambda_t(t-r_0)\nu_0) \) for some \( h(t, i) = \delta_{u_t(i)} \) for each \( i \in S \), where \( t \mapsto u_t(i) \) is a curve in \( U \) and \( \delta_x \) denote the Dirac measure in \( U \). Then \( \mu_t \) is usually called a deterministic policy, see [9] for more details.

In this paper we impose the following assumptions on the primitive \( Q \)-matrix of the continuous-time Markov decision process \( (\Lambda_t) \).

**Assumptions:**

(H1) \( \mu \mapsto q_{ij}(\mu) \) is continuous for every \( i, j \in S \), and \( M := \sup_{i \in S} \sup_{\mu \in \mathcal{P}(U)} q_i(\mu) < \infty \).

(H2) There exists a compact function \( \Phi : S \to [1, \infty) \), a compact set \( B_0 \subset S \), constants \( \lambda_0 > 0 \) and \( \kappa_0 \geq 0 \) such that

\[
Q_\mu \Phi(i) := \sum_{j \neq i} q_{ij}(\mu)(\Phi(j) - \Phi(i)) \leq \lambda_0 \Phi(i) + \kappa_0 1_{B_0}(i).
\]

(H3) There exists a \( K \in \mathbb{N} \) such that for every \( i \in S \) and \( \mu \in \mathcal{P}(U) \), \( q_{ij}(\mu) = 0 \), if \( |j-i| > K \).

Here if for every \( c \in \mathbb{R} \), the set \( \{ i \in S ; \Phi(i) \leq c \} \) is a compact set, then \( \Phi \) is called a compact function. Condition (H3) is a technical condition, which is used when we consider to use the dominated convergence theorem in the argument of our main theorem.

In contrast to the well-studied continuous-time Markov decision process, the controlled system \( (\Lambda_t) \) studied in this work is no longer a Markov chain, and the delay-dependent control policy makes it more difficult to describe the distribution of \( (\Lambda_t) \). In [25], Shao and Zhao considered the optimal control problem with the consideration of the random impact of the environment. By developing the classical compactness method to the optimal control problem for solutions of stochastic differential equations (cf. e.g. Kushner [16], Haussmann and Suo [12, 13] and references therein), they succeeded in solving the optimal control problem for Markov decision processes in random environments by noting that the control policies for Markov decision processes are corresponding to some kind of feedback controls with the restriction on remaining Markovian property. In this work we shall show that this method is still valid to deal with the optimal delay-dependent control problem for continuous-time Markov decision processes.

Let

\[
\mathcal{U} = \{ \mu : [0, T] \to \mathcal{P}(U) \text{ is measurable} \}.
\] (2.6)
\( \mathcal{U} \) can be viewed as a subspace of \( \mathcal{P}([0,T] \times U) \) through the map

\[
(\mu_t)_{t \in [0,T]} \mapsto \bar{\mu},
\]

where \( \bar{\mu} \) is determined by

\[
\bar{\mu}(A \times B) = \frac{1}{T} \int_A \mu_t(B) dt.
\]

Endow \( \mathcal{U} \) the induced weak convergence topology from \( \mathcal{P}([0,T] \times U) \). This topology is equivalent to the topology induced by the following Wasserstein distance on \( \mathcal{P}([0,T] \times U) \):

\[
W_1(\bar{\mu}, \bar{\nu}) = \inf_{\Gamma \in \mathcal{C}(\bar{\mu}, \bar{\nu})} \int_{([0,T] \times U)^2} (|s-t| + |x-y|) d\Gamma((s,x), (t,y)),
\]

where \( \mathcal{C}(\bar{\mu}, \bar{\nu}) \) stands for the collection of couplings of \( \bar{\mu} \) and \( \bar{\nu} \) over \( ([0,T] \times U)^2 \). The canonical path space for our problem is defined as

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

endowed with the product topology, which is a metrizable and separable space (cf. [12]). Denote by \( \hat{\mathcal{D}}^1 \) (resp. \( \hat{\mathcal{D}}^2 \)) the Borel \( \sigma \)-algebra of \( \mathcal{D}([0,T]; \mathcal{S}) \) (resp. \( \mathcal{U} \)), and \( \hat{\mathcal{D}}^1_t \) (resp. \( \hat{\mathcal{D}}^2_t \)) the \( \sigma \)-algebra up to time \( t \). Define the \( \sigma \)-algebra of \( \hat{\Omega} \) as

\[
\hat{\mathcal{F}} := \hat{\mathcal{D}}^1 \times \hat{\mathcal{D}}^2, \quad \text{and} \quad \hat{\mathcal{F}}_t = \hat{\mathcal{D}}^1_t \times \hat{\mathcal{D}}^2_t.
\]

For each delay control \( \alpha = (\Lambda_t, \mu_t, x, i) \in \Pi_{s,i} \), we define a measurable map \( \Phi_\alpha : \Omega \to \hat{\Omega} \) as

\[
\Phi_\alpha(\omega) = (\Lambda_t(\omega), \mu_t(\omega))_{t \in [0,T]}, \quad \Lambda_r(\omega) \equiv i, \ \mu_r(\omega) \equiv \mu_s, \ 0 \leq r \leq s.
\]

Then, there exists a corresponding probability on \( (\hat{\Omega}, \hat{\mathcal{F}}) \) defined by \( R = \mathbb{P} \circ \Phi_\alpha^{-1} \). We denote by \( \hat{\Pi}_{s,i} \) the space of probabilities induced by the delay-dependent control set \( \Pi_{s,i} \) with initial condition \( (s,i) \). By the definition of value function, we have

\[
V(s,i) = \inf_{\alpha \in \Pi_{s,i}} J(s,i,\alpha) = \inf_{R \in \hat{\Pi}_{s,i}} \mathbb{E}_R \left[ \int_s^T f(t, \Lambda_t, \mu_t) dt + g(\Lambda_T) \right].
\]

The topology and properties of the canonical path space have been well studied, see, for instance [12, 18, 27], and the references therein.
3 Existence of optimal delay-dependent controls

By developing the compactness method presented for instance in [12] and [16], Shao [26] investigated the optimal control problem for the regime-switching processes. There, the control on the transition rate matrix of the jumping process \((\Lambda_t)\) has been studied. In this paper we shall apply the result [26, Theorem 2.3] to the current situation to obtain the existence of optimal delay-dependent controls of our continuous-time Markov decision processes under the mild conditions (H1)–(H3).

**Theorem 3.1** Assume (H1)-(H3) hold. Then for every \(s \in [0, T)\), \(i \in S\), there exists an optimal delay-dependent control \(\alpha^* \in \Pi_{s,i}\).

**Proof.** This theorem is proved by using the idea of [26, Theorem 2.3]. The proof is a little long. In order to save space, here we only sketch the idea and point out the different points compared with that of [26, Theorem 2.3].

We only need to consider the nontrivial case \(V(s,i) < \infty\). For simplicity of notation, we consider the case \(s = 0\), and separate the proof into three steps.

**Step 1.** According to the definition of \(V(0,i)\), there exists a sequence of delay-dependent controls \(\alpha_n = (\Lambda_t^{(n)}, \mu_t^{(n)}, 0, i) \in \Pi_{0,i}\) such that
\[
\lim_{n \to \infty} J(0, i, \alpha_n) = V(0, i).
\] (3.1)

Denote by \(R_n\) the probability measures on \((\hat{\Omega}, \hat{\mathcal{F}})\) corresponding to \(\alpha_n\). Let \(\mathcal{L}_\mu^n\) (resp. \(\mathcal{L}_\Lambda^n\)) be the marginal distribution of \(R_n\) with respect to \((\mu_t^{(n)})_{t \in [0,T]}\) (resp. \((\Lambda_t^{(n)})_{t \in [0,T]}\)) in \(\mathcal{U}\) (resp. \(\mathcal{D}([0,T]; S)\)). Since \(\mathcal{P}([0,T] \times U)\) is compact and further \(\mathcal{U}\) is compact as a closed subset, we have \((\mathcal{L}_\mu^n)_{n \geq 1}\) is tight.

We proceed to prove that \((\mathcal{L}_\Lambda^n)_{n \geq 1}\) is tight. For each \(n \geq 1\), by (H2) and Itô-Dynkin’s formula, we have
\[
\mathbb{E}\Phi(\Lambda_t^{(n)}) = \Phi(i) + \mathbb{E} \int_0^t Q_{t,s} \Phi(\Lambda_s^{(n)}) ds \\
\leq \Phi(i) + \mathbb{E} \int_0^t (\lambda_0 \Phi(\Lambda_s^{(n)}) + \kappa_0) ds,
\]
which yields from Gronwall’s inequality that
\[
\mathbb{E}\Phi(\Lambda_t^{(n)}) \leq (\Phi(i) + \kappa_0 T) e^{\lambda_0 t}, \quad t \in [0, T], \ n \geq 1.
\] (3.2)
For any $\varepsilon > 0$, we can find $N_\varepsilon > 0$ such that
\[
\sup_n \mathbb{P}(\Lambda_t^{(n)} \in K_\varepsilon) \leq \sup_n \frac{\mathbb{E}\Phi(\Lambda_t^{(n)})}{N_\varepsilon} \leq \frac{(\Phi(i) + \kappa_0 T)e^{\lambda_0 T}}{N_\varepsilon} < \varepsilon, \tag{3.3}
\]
where $K_\varepsilon = \{ j \in \mathcal{S}; \Phi(j) \leq N_\varepsilon \}$. Since $\Phi$ is a compact function, $K_\varepsilon$ is a compact set. Moreover, for every $0 \leq u \leq \delta$, due to (H1),
\[
\mathbb{E}[1_{\Lambda_t^{(n)} \neq \Lambda_t^{(n)}_{s u} = \Lambda_t^{(n)}}] \leq 1 - P(\Lambda_s^{(n)} = \Lambda_t^{(n)}) \leq 1 - e^{-Mu} \leq 1 - e^{-M\delta} =: \gamma_n(\delta). \tag{3.4}
\]
To apply [6, Theorem 8.6, p.138], by taking $q(i,j) = 1_{i \neq j}$, $\beta = 1$, and $\gamma_n(\delta)$ given in (3.4), and invoking (3.3), we obtain the that tightness of $(L^n_t)_{n \geq 1}$.

**Step 2.** Since the marginal distributions $(L^n_t)_{n \geq 1}$ and $(M^n_t)_{n \geq 1}$ are both tight, $(R^*_n)_{n \geq 1}$ is tight as well. Hence, there exists a subsequence $n_k$, $k \geq 1$, such that $R^*_{n_k}$ weakly converges to some probability measure $R_0$ on $(\hat{\Omega}, \hat{\mathcal{F}})$ as $k \to \infty$. By virtue of Skorokhod’s representation theorem (cf. e.g. [6, Chapter 3]), there exists a probability space $(\Omega', \mathcal{F}', P')$ on which are defined a sequence of $\hat{\Omega}$-valued random variables $Y_{n_k} = (\Lambda_t^{(n_k)}, \mu_t^{(n_k)})_{t \in [0,T]}$ with distribution $R_{n_k}$, $k \geq 1$, and $Y_0 = (\Lambda_t^{(0)}, \mu_t^{(0)})_{t \in [0,T]}$ with distribution $R_0$ such that
\[
\lim_{k \to \infty} Y_{n_k} = Y_0, \quad P'-a.s. \tag{3.5}
\]
Analogous to the Step 2 in the argument of [26, Theorem 2.3], we can show that $\alpha^* := (\Lambda_t^{(0)}, \mu_t^{(0)}, 0, i)$ is a delay-dependent control in $\Pi_{0,i}$. During this procedure, we need to replace the sigma fields $\mathcal{F}^\Lambda_{-n,t}$ by the following
\[
\mathcal{F}^\Lambda_{-n,t} := \sigma\{(\Lambda_t^{(k)}, \ldots, \Lambda_t^{(k-mn)})_{k \geq n}\}.
\]
**Step 3.** Invoking (3.1) and the lower semi-continuity of $f$ and $g$, we obtain
\[
V(0, i) = \lim_{k \to \infty} \mathbb{E}\left[\int_0^T f(t, \Lambda_t^{(n_k)}, \mu_t^{(n_k)})dt + g(\Lambda_T^{(n_k)})\right] \\
\geq \mathbb{E}\left[\int_0^T f(t, \Lambda_t^{(0)}, \mu_t^{(0)})dt + g(\Lambda_T^{(0)})\right] \\
\geq V(0, i).
\]
By taking $\alpha^* = (\Lambda_t^{(0)}, \mu_t^{(0)}, 0, i) \in \Pi_{0,i}$, the previous inequalities imply that $\alpha^*$ is an optimal delay-dependent control of the continuous-time Markov jump process. The proof of this theorem is complete.\]
4 Dynamic programming principle and viscosity solution.

In the rest of the paper, we introduce the dynamic programming principle for the controlled processes with delay-dependent control and the differential equation corresponding to the value function. To do so, we give some notations. Assume that $\tau$ is an $\hat{\mathcal{F}}_t$-stopping time satisfying $0 \leq \tau \leq T$, $\hat{\mathcal{F}}_\tau$ is denoted by the collection of sets $A$ such that $A \cap \{\tau \leq t\} \in \hat{\mathcal{F}}_t$, $\forall t \in [0, T]$.

**Theorem 4.1** Assume (H1)-(H3) hold. For each $\hat{\mathcal{F}}_\tau$-stopping time $\tau$ satisfying $s \leq \tau \leq T$, then

$$V(s, i) = \inf \left\{ \mathbb{E}_R \left[ \int_s^\tau f(t, \Lambda_t, \mu_t)dt + V(\tau, \Lambda_\tau) \right] ; R \in \hat{\Pi}_{s,i} \right\}.$$

**Proof.** Define a subset of $\hat{\Pi}_{s,i}$ as

$$\hat{\Pi}_{s,i}^0 = \left\{ R \in \hat{\Pi}_{s,i} : V(s, i) = \mathbb{E}_R \left[ \int_s^T f(t, \Lambda_t, \mu_t)dt + g(\Lambda_T) \right] \right\}.$$

By Theorem 3.1, $\hat{\Pi}_{s,i}^0 \neq \emptyset$ for any $s \in [0, T]$ and $i \in \mathcal{S}$. According to measurable choices theorem presented by Stroock and Varadhan [27], there exists a Borel-measurable map $H : [0, t] \times \mathcal{S} \to \mathcal{P}(U)$, which is called measurable selector, satisfying for each $(s, i) \in [0, t] \times \mathcal{S}$, $H(s, i) \in \hat{\Pi}_{s,i}^0$. Refer to [12, Lemma 3.9] or [25, Proposition 4.2] for more details of the existence of the measurable selector. Hence, for any $\hat{\omega} \in \hat{\Omega}$, $H(\tau(\hat{\omega}), \Lambda_\tau(\hat{\omega}))$ is a probability measure on $(\hat{\Omega}, \hat{\mathcal{F}})$ and satisfies

$$V(\tau(\hat{\omega}), \Lambda_\tau(\hat{\omega})) = \mathbb{E}_{H(\tau(\hat{\omega}), \Lambda_\tau(\hat{\omega}))} \left[ \int_{\tau(\hat{\omega})}^T f(t, \Lambda_t, \mu_t)dt + g(\Lambda_T) \right]. \tag{4.1}$$

Note that the topology on $\hat{\Omega}$ is separable, then $\hat{\mathcal{F}}_t$ is countably generated, and then for every probability measure $\mathbb{P}$ on $(\hat{\Omega}, \hat{\mathcal{F}})$, the regular conditional probability distribution of $\mathbb{P}$ for given $\mathcal{F}_\tau$ exists, cf. [12, 13]. According to [13, Lemma 3.3], for each $R \in \hat{\Pi}_{s,i}$, there exists a unique probability measure, denoted by $R^H$, such that $R^H(A) = R(A)$, $\forall \ A \in \mathcal{F}_\tau$ and the regular conditional probability distribution of $R^H$ for given $\mathcal{F}_\tau$ is $H(\cdot, \Lambda_{\tau(\cdot)})$. Moreover, by [13, Proposition 3.8], it holds that $R^H \in \hat{\Pi}_{s,i}$. Hence, we have

$$V(\tau(\hat{\omega}), \Lambda_\tau(\hat{\omega})) = \mathbb{E}_{R^H} \left[ \int_{\tau(\hat{\omega})}^T f(t, \Lambda_t, \mu_t)dt + g(\Lambda_T) \bigg| \mathcal{F}_\tau \right].$$
Due to the definition of value function $V(s, i)$, we have

$$V(s, i) \leq \mathbb{E}_{R^H} \left[ \int_s^\tau f(t, \Lambda_t, \mu_t)dt + \int_\tau^T f(t, \Lambda_t, \mu_t)dt + g(\Lambda_T) \right]$$

$$= \mathbb{E}_{R^H} \left[ \int_s^\tau f(t, \Lambda_t, \mu_t)dt + \mathbb{E}_{R^H} \left[ \int_\tau^T f(t, \Lambda_t, \mu_t)dt + g(\Lambda_T) \mid \mathcal{F}_\tau \right] \right]$$

$$= \mathbb{E}_R \left[ \int_s^\tau f(t, \Lambda_t, \mu_t)dt + V(\tau, \Lambda_\tau) \right],$$

where the last equation is based on the relationship between $R$ and $R^H$. The arbitrariness of $R \in \hat{\Pi}_{s,i}$ implies that

$$V(s, i) \leq \inf \left\{ \mathbb{E}_R \left[ \int_s^\tau f(t, \Lambda_t, \mu_t)dt + V(\tau, \Lambda_\tau) \right] ; R \in \hat{\Pi}_{s,i} \right\}.$$

Conversely, by Theorem 3.1, there exists an optimal delay-dependent control $\alpha^* \in \Pi_{s,i}$ and then denote by $R^* \in \hat{\Pi}_{s,i}$ the corresponding probability measure on $(\hat{\Omega}, \hat{\mathcal{F}})$. Then we have

$$V(s, i) = \mathbb{E}_{R^*} \left[ \int_s^\tau f(t, \Lambda_t, \mu_t)dt + \int_\tau^T f(t, \Lambda_t, \mu_t)dt + g(\Lambda_T) \right]$$

$$\geq \mathbb{E}_{R^*} \left[ \int_s^\tau f(t, \Lambda_t, \mu_t)dt + V(\tau, \Lambda_\tau) \right]$$

$$\geq \inf \left\{ \mathbb{E}_R \left[ \int_s^\tau f(t, \Lambda_t, \mu_t)dt + V(\tau, \Lambda_\tau) \right] ; R \in \hat{\Pi}_{s,i} \right\}.$$

The dynamic programming principle is thus proved. \(\Box\)

The next result is about the continuity of value function. Since $S$ is a countable state space equipped with discrete topology, we only need to consider the continuity of $V(s, i)$ in the time variable $s$.

**Proposition 4.2** Assume (H1)-(H3) hold. Support that $f, g$ are bounded and $f$ satisfies the following condition,

$$|f(t, i, \mu) - f(s, i, \mu)| \leq C_0|t - s|, \quad 0 \leq s, t \leq T, \quad (4.2)$$

uniformly for $i \in S$ and $\mu \in \mathcal{P}(U)$. Then, the value function $V(s, i)$ is Lipschitz continuous with respect to the time variable $s$. In fact, fixed any $i \in S$, there exists a constant $C > 0$ such that

$$|V(s, i) - V(s', i)| \leq C|s - s'|, \quad 0 \leq s, s' \leq T.$$
Proof. For convenience, denote by $C_1$ and $C_2$ the constants such that

$$\sup_{(t,i,\mu) \in [0,T] \times S \times \mathcal{P}(U)} |f(t,i,\mu)| \leq C_1 \quad \text{and} \quad \sup_{i \in S} |g(i)| \leq C_2.$$ 

Fix any $i \in S$ and assume $0 \leq s \leq s' \leq T$. According to Theorem 3.1, there exists an optimal delay-dependent control $\alpha^* = (\Lambda_t, \mu_t, s, i) \in \Pi_{s,i}$ such that $V(s,i) = J(s,i,\alpha^*)$. By time shift, we can define a process with the initial point $(s',i)$ as following

$$\Lambda'_t = \Lambda_t - \Delta s, \quad \mu'_t = \mu_t - \Delta s, \quad \forall t \in [s',T],$$

where $\Delta s := s' - s$. It is easy to verify that (2.1) and (2.3) hold for $(\Lambda'_t, \mu'_t)$, which means that $\alpha' := (\Lambda'_t, \mu'_t, s', i)$ is a delay-dependent control in $\Pi_{s',i}$. Using (H1) and (2.1), we have

$$\mathbb{E} \left[ 1_{\Lambda'_t \neq \Lambda_t} \right] = \mathbb{P} (\Lambda_t - \Delta s \neq \Lambda_t) \leq M\Delta s + o(\Delta s).$$

Using (4.2), we obtain

$$\mathbb{E} \left[ \int_{s'}^{s} |f(t, \Lambda_t, \mu_t)| dt \right] \leq C_1 \Delta s, \quad \text{and}$$

$$\mathbb{E} \left[ |g(\Lambda_T') - g(\Lambda_T)| \right] \leq 2C_2 \mathbb{E} \left[ 1_{\Lambda'_T \neq \Lambda_T} \right] \leq 2MC_2 \Delta s + o(\Delta s).$$

To estimate the first term of (4.3), we combine the boundedness and (4.2),

$$\mathbb{E} \left[ \int_{s'}^{T} |f(t, \Lambda'_t, \mu'_t) - f(t, \Lambda_t, \mu_t)| dt \right]$$

$$= \mathbb{E} \left[ \int_{s}^{T-\Delta s} |f(t+\Delta s, \Lambda_t, \mu_t)dt - f(t, \Lambda_t, \mu_t)| dt \right] + \mathbb{E} \left[ \int_{s}^{s'} |f(t, \Lambda_t, \mu_t)| dt \right]$$

$$+ \mathbb{E} \left[ \int_{T-\Delta s}^{T} |f(t, \Lambda_t, \mu_t)| dt \right]$$

$$\leq TC_0 \Delta s + 2C_1 \Delta s.$$
Hence,

$$|V(s, i) - V(s', i)| \leq (3C_1 + 2MC_2 + TC_0)\Delta s + o(\Delta s).$$

By the symmetric position of $s$ and $s'$, we have $|V(s, i) - V(s', i)| \leq C|s - s'|$. □

According to Proposition 4.2 and Ramemacher’s theorem, we know that $t \mapsto V(t, i)$ is almost everywhere differentiable in $[0, T]$ with respect to Lebesgue measure. But, it is not easy to justify whether $V(t, i)$ is differentiable everywhere in $[0, T]$. In such situation, we need to introduce the concept of viscosity solution to further characterize $V(t, i)$. Consider the following equation

$$-\frac{\partial v}{\partial t} - \inf_{\mu \in \mathcal{P}(U)} \left\{ \sum_{j \neq i} q_{ij}(\mu)(v(t, j) - v(t, i)) + f(t, i, \mu) \right\} = 0. \quad (4.4)$$

**Definition 4.3** Let $v : [0, T] \times S \rightarrow \mathbb{R}$ be a continuous function.

(i) $v$ is called a viscosity subsolution of (4.4) if

$$-\frac{\partial \phi}{\partial t}(t_0, i_0) - \inf_{\mu \in \mathcal{P}(U)} \left\{ \sum_{j \neq i_0} q_{i_0j}(\mu)(\phi(t_0, j) - \phi(t_0, i_0)) + f(t_0, i_0, \mu) \right\} \leq 0$$

for all $(t_0, i_0) \in [0, T] \times S$ and for all $\phi \in C^1([0, T] \times S)$ such that $(t_0, i_0)$ is a minimum point of $v - \phi$.

(ii) $v$ is called a viscosity supersolution of (4.4) if

$$-\frac{\partial \phi}{\partial t}(t_0, i_0) - \inf_{\mu \in \mathcal{P}(U)} \left\{ \sum_{j \neq i_0} q_{i_0j}(\mu)(\phi(t_0, j) - \phi(t_0, i_0)) + f(t_0, i_0, \mu) \right\} \geq 0$$

for all $(t_0, i_0) \in [0, T] \times S$ and for all $\phi \in C^1([0, T] \times S)$ such that $(t_0, i_0)$ is a maximum point of $v - \phi$.

(iii) $v$ is called a viscosity solution of (4.4) if it is both a viscosity subsolution and a viscosity supersolution of (4.4).

The next result says that the value function is a solution to the Hamilton-Jacobi-Bellman equation (4.4) in the viscosity sense.

**Theorem 4.4** Under the conditions of Proposition 4.2, the value function $V(t, i)$ is a viscosity solution of the equation (4.4).
Proof. We first consider the viscosity supersolution property. Let \((t_0, i_0) \in [0, T) \times \mathcal{S}\) and \(\phi \in C^1([0, T) \times \mathcal{S})\) be a test function such that

\[
0 = (V - \phi)(t_0, i_0) = \max \{(V - \phi)(t, i); (t, i) \in [0, T) \times \mathcal{S}\}. \tag{4.5}
\]

Take an arbitrary point \(\tilde{\mu} \in \mathcal{P}(U)\), and let

\[
\mu_t = \tilde{\mu}, \quad \forall t \in [s, T],
\]

which is a constant control policy and obviously satisfies the conditions of Definition 2.1. According to the dynamic programming principle (Theorem 4.1), we have

\[
V(t_0, i_0) \leq E \left[ \int_{t_0}^{t} f(r, \Lambda_r, \tilde{\mu})dr + V(t, \Lambda_t) \right].
\]

Due to (4.5), it holds \(V \leq \phi\), and hence

\[
\phi(t_0, i_0) \leq E \left[ \int_{t_0}^{t} f(r, \Lambda_r, \tilde{\mu})dr + \phi(t, \Lambda_t) \right]. \tag{4.6}
\]

Applying Itô-Dynkin’s formula to the function \(\phi\) (cf. [10, Theorem 3.1]), we get

\[
E \phi(t, \Lambda_t) = \phi(t_0, i_0) + E \left[ \int_{t_0}^{t} \left( \frac{\partial \phi}{\partial r}(r, \Lambda_r) + Q(\tilde{\mu})\phi(r, \Lambda_r) \right)dr \right]. \tag{4.7}
\]

Inserting (4.7) into (4.6) leads to

\[
- E \left[ \int_{t_0}^{t} \left( \frac{\partial \phi}{\partial r}(r, \Lambda_r) + Q(\tilde{\mu})\phi(r, \Lambda_r) + f(r, \Lambda_r, \mu_r) \right)dr \right] \leq 0. \tag{4.8}
\]

Dividing both sides of (4.8) by \(t - t_0\) and letting \(t \downarrow t_0\), we get from the almost sure right-continuity of the trajectories of \((\Lambda_t)\) that

\[
- \frac{\partial \phi}{\partial t}(t_0, i_0) - \sum_{j \neq i_0} q_{i_0 j}(\tilde{\mu})(\phi(t_0, j) - \phi(t_0, i_0)) + f(t_0, i_0, \tilde{\mu}) \leq 0. \tag{4.9}
\]

Then, by the arbitrariness of \(\tilde{\mu} \in \mathcal{P}(U)\), \(V(t, i)\) is a viscosity supersolution of (??).

Next, we proceed to the viscosity subsolution property. Let \((t_0, i_0) \in [0, T) \times \mathcal{S}\) and \(\phi \in C^1([0, T) \times \mathcal{S})\) be a test function such that

\[
0 = (V - \phi)(t_0, i_0) = \min \{(V - \phi)(t, i); (t, i) \in [0, T) \times \mathcal{S}\}. \tag{4.10}
\]
The desired result will be shown by contradiction. Assume
\[
- \frac{\partial \phi}{\partial t}(t_0, i_0) - \inf_{\mu \in \mathcal{P}(U)} \left\{ Q(\mu) \phi(t_0, i_0) + f(t_0, i_0, \mu) \right\} < 0. \tag{4.11}
\]
By (H1), the compactness of \( \mathcal{P}(U) \) and the continuity of \( f \), we obtain from (4.11) that there exist \( \varepsilon, \eta > 0 \) such that for any \( 0 \leq t - t_0 \leq \eta \), it holds
\[
- \frac{\partial \phi}{\partial t}(t, i_0) - \inf_{\mu \in \mathcal{P}(U)} \left\{ Q(\mu) \phi(t, i_0) + f(t, i_0, \mu) \right\} \leq -\varepsilon. \tag{4.12}
\]

Let \((t_k)_{k \geq 1}\) be a sequence satisfying \( \lim_{k \to \infty} t_k = t_0 \). Using the dynamic programming principle (Theorem 4.1) again, for each \( k \geq 1 \), there exists \( \alpha^{(k)} = (\Lambda^{(k)}_t, \mu^{(k)}_t, t_0, i_0) \in \Pi_{t_0, i_0} \) such that
\[
V(t_0, i_0) \geq \mathbb{E} \left[ \int_{t_0}^{\beta_k} f(r, \Lambda^{(k)}_r, \mu^{(k)}_r) dr + V(\beta_k, \Lambda^{(k)}_{\beta_k}) \right] - \frac{\varepsilon}{2}(t_k - t_0),
\]
where \( \beta_k = t_k \land \tau_k \), and \( \tau_k \) defined by
\[
\tau_k = \inf \{ t \in [t_0, T] ; \Lambda^{(k)}_t \neq \Lambda^{(k)}_{t_0} \} \land (t_0 + \eta). \tag{4.13}
\]
Due to (4.10), we have \( V \geq \phi \) and
\[
\phi(t_0, i_0) \geq \mathbb{E} \left[ \int_{t_0}^{\beta_k} f(r, \Lambda^{(k)}_r, \mu^{(k)}_r) dr + \phi(\beta_k, \Lambda^{(k)}_{\beta_k}) \right] - \frac{\varepsilon}{2}(t_k - t_0). \tag{4.14}
\]
Using Itô-Dynkin's formula to the function \( \phi \), we have
\[
\mathbb{E} \left[ \int_{t_0}^{\beta_k} f(r, \Lambda^{(k)}_r, \mu^{(k)}_r) + \frac{\partial \phi}{\partial r} + Q(\mu^{(k)}_r) \phi(r, \Lambda^{(k)}_r) dr \right] \leq \frac{\varepsilon}{2}(t_k - t_0).
\]
Then (4.12) and the definition of \( \beta_k \) implies that
\[
\frac{\mathbb{E}[\beta_k - t_0]}{t_k - t_0} \leq \frac{1}{2}, \quad k \geq 1. \tag{4.15}
\]
On the other hand, by (H1), we have
\[
\mathbb{P}(\beta_k - t_0 \leq t_k - t_0) \leq \mathbb{P} \left( \sup_{s \in [t_0, t_k]} |\Lambda^{(k)}_s - \Lambda^{(k)}_{t_0}| > 0 \right) \leq 1 - e^{-M(t_k - t_0)},
\]
Therefore,
\[
\lim_{k \to \infty} \mathbb{P}(\beta_k - t_0 \geq t_k - t_0) = 1.
\]
Since
\[ P(\beta_k - t_0 \geq t_k - t_0) \leq \frac{\mathbb{E}[\beta_k - t_0]}{t_k - t_0} \leq 1, \]
we get finally that
\[ \lim_{k \to \infty} \frac{\mathbb{E}[\beta_k - t_0]}{t_k - t_0} = 1, \tag{4.16} \]
which contradicts (4.15). Consequently, we have
\[ -\frac{\partial \phi}{\partial t}(t_0, i_0) - \inf_{\mu \in \mathcal{P}(U)} \left\{ \sum_{j \neq i_0} (\phi(t_0, j) - \phi(t_0, i_0)) + f(t_0, i_0, \mu) \right\} \geq 0. \tag{4.17} \]
This means that \( V(t, i) \) is a viscosity subsolution of (4.4). We conclude the proof of this theorem by the definition of viscosity solution of (4.4).

In the end let us discuss the uniqueness of the viscosity solution of (4.4). For this purpose it is sufficient to establish the following comparison principle for (4.4). We shall develop the method used to establish the comparison principle for Hamilton-Jacobi-Bellman equations associated with diffusion processes to the equations associated with purely jumping processes.

**Theorem 4.5** Assume the conditions of Proposition 4.2 hold. Let \( V_1 \) (resp. \( V_2 \)) be a viscosity subsolution (resp. supersolution) of (4.4) in \([0, T) \times \mathcal{S}\). Then
\[ \sup_{[0, T] \times \mathcal{S}} [V_2 - V_1] = \sup_{\{T\} \times \mathcal{S}} [V_2 - V_1]. \]

**Proof.** Obviously, we just need to show that
\[ \sup_{[0, T] \times \mathcal{S}} [V_2 - V_1] \leq \sup_{\{T\} \times \mathcal{S}} [V_2 - V_1]. \tag{4.18} \]
Given arbitrary \( i_0 \in \mathcal{S} \), according to Proposition 4.2, the continuity of \( V_1 \) and \( V_2 \) implies that \( K_{i_0} := \sup_{t \in [0, T]} |V_1(t, i_0)| \vee |V_2(t, i_0)| < \infty \). Define a function on \([0, T] \times [0, T]\) as
\[ \Psi_{i_0}(t, s) = V_2(t, i_0) - V_1(s, i_0) - \frac{1}{2\delta}(t - s)^2 - \frac{\beta}{\delta^2}(t \vee s - T), \]
where \( \delta, \beta > 0 \) are two parameters. Again, the continuities of \( V_1 \) and \( V_2 \) imply that \( \Psi_{i_0} \) achieves the maximum on \([0, T] \times [0, T]\). Denoted by \((\bar{t}, \bar{s}) \in [0, T] \times [0, T] \) an arbitrary one of the maximum points.
We first give an estimate of the distance between \( \bar{s} \) and \( \bar{t} \). For any \( \rho \geq 0 \), let

\[
D_\rho = \left\{ (t, s) \in [0, T] \times [0, T] : |t - s|^2 \leq \rho \right\},
\]

\[
m_{i_0}^{(1)}(\rho) = 2 \sup \left\{ |V_1(t, i_0) - V_1(s, i_0)| : (t, s) \in D_\rho \right\},
\]

\[
m_{i_0}^{(2)}(\rho) = 2 \sup \left\{ |V_2(t, i_0) - V_2(s, i_0)| : (t, s) \in D_\rho \right\}.
\]

Then \( m_{i_0}^{(1)} \) and \( m_{i_0}^{(2)} \) are increasing functions satisfying \( m_{i_0}^{(1)}(0) = m_{i_0}^{(2)}(0) = 0 \). Moreover, it follows from the continuity of \( V_1 \) and \( V_2 \) and the compactness of \([0, T] \times [0, T]\) that \( m_{i_0}^{(1)}, m_{i_0}^{(2)} \) are continuous. Since \( V_1(\cdot, i_0) \) and \( V_2(\cdot, i_0) \) are bounded, \( m_{i_0}^{(1)} \) and \( m_{i_0}^{(2)} \) are bounded as well and denoted by \( M_{i_0} := \sup \{m_{i_0}^{(1)}(\rho) \vee m_{i_0}^{(2)}(\rho) : \rho \geq 0\} < \infty \). Note that \( \Psi_{i_0}(\bar{t} \lor \bar{s}, \bar{t} \lor \bar{s}) \leq \Psi_{i_0}(\bar{t}, \bar{s}) \leq \sqrt{\delta M_{i_0}} \).

Hence, we have

\[
|\bar{t} - \bar{s}| \leq \sqrt{\delta M_{i_0}}. \tag{4.19} \text{supp1}
\]

If \( \bar{s} = T \), then it holds that

\[
\Psi_{i_0}(\bar{t}, \bar{s}) \leq V_2(\bar{t}, i_0) - V_1(\bar{s}, i_0) \leq V_2(\bar{t}, i_0) - V_2(\bar{s}, i_0) + \sup_{(T) \times S} [V_2 - V_1].
\]

By the definition of \( m_{i_0}^{(2)} \) and (4.19), we obtain

\[
\Psi_{i_0}(\bar{t}, \bar{s}) \leq \frac{1}{2} m_{i_0}^{(2)}(\delta M_{i_0}) + \sup_{(T) \times S} [V_2 - V_1]. \tag{4.20} \text{supp2}
\]

Similarly, if \( \bar{t} = T \), we have

\[
\Psi_{i_0}(\bar{t}, \bar{s}) \leq \frac{1}{2} m_{i_0}^{(1)}(\delta M_{i_0}) + \sup_{(T) \times S} [V_2 - V_1]. \tag{4.21} \text{supp3}
\]

Next, we shall show by contradiction that at least one of \( \bar{s} \) and \( \bar{t} \) equals to \( T \). Assume that both \( \bar{t} \) and \( \bar{s} \) are in \([0, T]\). Define an auxiliary function on \([0, T] \times S\) as

\[
\psi_{i_0}^{(1)}(s, j) = -\frac{1}{2\delta}(\bar{t} - s)^2 - \left( 2K_{i_0} + \frac{T^2}{\delta} + \frac{\beta T}{\delta^2} \right)(1 - 1_{i_0}(j)) - \frac{\beta}{\delta^2}(\bar{t} \lor s - T).
\]
For each $s \in [0,T]$, since $\Psi_{i_0}(\bar{t},s) \leq \Psi_{i_0}(\bar{t},\bar{s})$, it holds that
\[
V_1(\bar{s},i_0) + \frac{1}{2\delta}(\bar{t} - \bar{s})^2 + \frac{\beta}{\delta^2}(\bar{t} \vee \bar{s} - T) \leq V_1(s,i_0) + \frac{1}{2\delta}(\bar{t} - s)^2 + \frac{\beta}{\delta^2}(\bar{t} \vee s - T),
\]
and for each $j \in S$ with $j \neq i_0$,
\[
2K_{i_0} + \frac{T^2}{\delta} + \frac{\beta T}{\delta^2} \geq V_1(\bar{s},i_0) - V_1(s,j) + \frac{1}{2\delta}(\bar{t} - s)^2 - \frac{1}{2\delta}(\bar{t} - s)^2 + \frac{\beta}{\delta^2}(\bar{t} \vee \bar{s} - \bar{t} \vee s).
\]
Hence, $(\bar{s},i_0)$ attains the minimum point of $V_1 - \psi_{i_0}^{(1)}$. Since $V_1$ is the viscosity subsolution of (4.4), we have
\[
-\frac{1}{\delta}(\bar{t} - \bar{s}) + \frac{\beta}{\delta^2}1_{[\bar{t},T]}(\bar{s}) - \inf_{\mu \in \mathcal{P}(U)} \left\{ -\left(2K_{i_0} + \frac{T^2}{\delta} + \frac{\beta T}{\delta^2}\right) q_{i_0}(\mu) + f(\bar{s},i_0,\mu) \right\} \leq 0. \tag{4.22} \text{ supp4}
\]
Similarly, consider the test function on $[0,T] \times S$ as
\[
\psi_{i_0}^{(2)}(t,j) = \frac{1}{2\delta}(t - s)^2 + \left(2K_{i_0} + \frac{T^2}{\delta} + \frac{\beta T}{\delta^2}\right) (1 - 1_{i_0}(j)) + \frac{\beta}{\delta^2}(t \vee \bar{s} - T).
\]
Then, the same arguments imply that for each $t \in [0,T]$,
\[
V_2(t,i_0) - \frac{1}{2\delta}(t - s)^2 - \frac{\beta}{\delta^2}(t \vee \bar{s} - T) \leq V_2(\bar{t},i_0) - \frac{1}{2\delta}(\bar{t} - s)^2 - \frac{\beta}{\delta^2}(\bar{t} \vee \bar{s} - T),
\]
and for each $j \in S$ with $j \neq i_0$,
\[
2K_{i_0} + \frac{T^2}{\delta} + \frac{\beta T}{\delta^2} \geq V_2(t,j) - V_2(\bar{t},i_0) + \frac{1}{2\delta}(\bar{t} - s)^2 - \frac{1}{2\delta}(\bar{t} - s)^2 + \frac{\beta}{\delta^2}(\bar{t} \vee \bar{s} - t \vee s),
\]
which mean that $(\bar{t},i_0)$ attains the maximum point of $V_2 - \psi_{i_0}^{(2)}$. Since $V_2$ is the viscosity supersolution of (4.4), we have
\[
-\frac{\beta}{\delta^2}1_{[\bar{t},T]}(\bar{t}) - \frac{1}{\delta}(\bar{t} - \bar{s}) + \inf_{\mu \in \mathcal{P}(U)} \left\{ \left(2K_{i_0} + \frac{T^2}{\delta} + \frac{\beta T}{\delta^2}\right) q_{i_0}(\mu) + f(\bar{t},i_0,\mu) \right\} \geq 0. \tag{4.23} \text{ supp5}
\]
If $\bar{t} < \bar{s}$, we have $1_{[\bar{t},T]}(\bar{s}) = 1$ and $1_{[\bar{t},T]}(\bar{t}) = 0$. Combining the inequalities (4.19), (4.22), (4.23) and (4.2), we arrive at
\[
\frac{\beta}{\delta^2} \leq \frac{1}{\delta}(\bar{t} - \bar{s}) + \inf_{\mu \in \mathcal{P}(U)} \left\{ -\left(2K_{i_0} + \frac{T^2}{\delta} + \frac{\beta T}{\delta^2}\right) q_{i_0}(\mu) + f(\bar{s},i_0,\mu) \right\}
\leq \inf_{\mu \in \mathcal{P}(U)} \left\{ -\left(2K_{i_0} + \frac{T^2}{\delta} + \frac{\beta T}{\delta^2}\right) q_{i_0}(\mu) + f(\bar{s},i_0,\mu) \right\}
\]
\[
- \inf_{\mu \in \mathcal{P}(U)} \left\{ \left( 2K_{i_0} + \frac{T^2}{\delta} + \frac{\beta T}{\delta^2} \right) q_{i_0}(\mu) + f(\bar{t}, i_0, \mu) \right\} \\
\leq \inf_{\mu \in \mathcal{P}(U)} \left\{ - \left( 2K_{i_0} + \frac{T^2}{\delta} \right) q_{i_0}(\mu) + f(\bar{s}, i_0, \mu) \right\} \\
- \inf_{\mu \in \mathcal{P}(U)} \left\{ \left( 2K_{i_0} + \frac{T^2}{\delta} \right) q_{i_0}(\mu) + f(\bar{t}, i_0, \mu) \right\} - \frac{\beta T}{\delta^2} \inf_{\mu \in \mathcal{P}(U)} \{ q_{i_0}(\mu) \}. \tag{4.24} \]
\]

If \( \bar{s} < \bar{t} \), we have \( 1_{[\bar{t}, T]}(\bar{s}) = 0 \) and \( 1_{(\bar{s}, T]}(\bar{t}) = 1 \), and the same method also implies (4.24).

Hence, we have

\[
\beta + \beta T \inf_{\mu \in \mathcal{P}(U)} q_{i_0}(\mu) \leq (4K_{i_0}\delta^2 + 2T^2\delta) \sup_{\mu \in \mathcal{P}(U)} q_{i_0}(\mu) + \delta^2 \sup_{\mu \in \mathcal{P}(U)} |f(\bar{s}, i_0, \mu) - f(\bar{t}, i_0, \mu)| \\
\leq (4K_{i_0}\delta^2 + 2T^2\delta) \sup_{\mu \in \mathcal{P}(U)} q_{i_0}(\mu) + \delta^2 C_0 \sqrt{\delta M_{i_0}}. \tag{4.25} \]

The right-hand side of (4.25) goes to 0 as \( \delta \downarrow 0 \), which leads to a contradiction with that \( \beta > 0 \). If \( \bar{t} = \bar{s} \), which means \( 1_{[\bar{t}, T]}(\bar{s}) = 1_{(\bar{s}, T]}(\bar{t}) = 1 \), the inequalities (4.22) and (4.23) imply that

\[
\frac{\beta}{\delta^2} \leq \frac{1}{2} \inf_{\mu \in \mathcal{P}(U)} \left\{ - \left( 2K_{i_0} + \frac{T^2}{\delta} + \frac{\beta T}{\delta^2} \right) q_{i_0}(\mu) + f(\bar{s}, i_0, \mu) \right\} \\
- \frac{1}{2} \inf_{\mu \in \mathcal{P}(U)} \left\{ \left( 2K_{i_0} + \frac{T^2}{\delta} + \frac{\beta T}{\delta^2} \right) q_{i_0}(\mu) + f(\bar{t}, i_0, \mu) \right\}.
\]

It still leads to contradiction with \( \beta > 0 \) as \( \delta \downarrow 0 \). Therefore we have either \( \bar{t} = T \) or \( \bar{s} = T \) or both. According to (4.20) and (4.21), we have

\[
\limsup_{\delta \downarrow 0} \Psi_{i_0}(\bar{t}, \bar{s}) \leq \sup_{(t, i_0) \in [0, T] \times \mathcal{S}} [V_2 - V_1]. \tag{4.26} \]

By the arbitrariness of \( i_0 \in \mathcal{S} \), we have

\[
V_2(t, i_0) - V_1(t, i_0) \leq \Psi_{i_0}(t, t) \leq \Psi_{i_0}(\bar{t}, \bar{s}), \quad \forall (t, i_0) \in [0, T] \times \mathcal{S}, \tag{4.27} \]

then (4.26) and (4.27) imply (4.18) after taking the limit \( \delta \downarrow 0 \). \( \Box \)

**Corollary 4.6** Under the conditions of Proposition 4.2, the value function \( V(t, i) \) is an unique viscosity solution of the equation (4.4).
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] N. Baüerle, U. Rieder, Markov Decision Processes with Applications to Finance, Springer, Heidelberg, 2011.

[3] P. Billinsley, Convergence of probability measures, John Wiley, New York, 1968.

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

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

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

[7] W.H. Fleming, H.M. Soner, Controlled Markov processes and viscosity solutions, Springer-Verlag, New York, 1975.

[8] M.K. Ghosh, S. Saha, Continuous-time controlled jump Markov processes on the finite horizon, In Optimization, Control, and Applications of Stochastic Systems. Birkhäuser, NewYork, (2012), 99-109.

[9] X.P. Guo, O. Hernández-Lerma, Continuous-time Markov decision processes. Theory and applications, Springer-Verlag, Berlin, 2009.

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

[11] X.P. Guo, Y.H. Huang, X.Y. Song, Linear programming and constrained average optimality for general continuous-time Markov decision processes in history-dependent policies, SIAM J. Control Optim. 50 (2012), 23-47.

[12] U.G. Haussmann, W. Suo, Singular optimal stochastic controls I: existence, SIAM J. Control Optim. 33 (1995), 916-936.
[13] U.G. Haussmann, W. Suo, Singular optimal stochastic controls II: dynamic programming, SIAM J. Control Optim. 33 (1995), 937-959.

[14] Y.H. Huang, Finite horizon continuous-time Markov decision processes with mean and variance criteria, Discrete Event Dynamic Systems. 28 (2018), 539-564.

[15] N. Krylov, Controllend diffusion processes, Springer-Verlag, New York, 1980.

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

[17] N. Li, Z. Wu, Maximum principle for anticipated recursive stochastic optimal control problem with delay and Lévy processes, Applied Mathematics-A Journal of Chinese Universities, 29(1) (2014), 67-85.

[18] P.A. Meyer, W.A. Zheng, Tightness criteria for laws of semimartingales, Ann. Inst. Henri Poincaré Probab. Statist., 20 (1984), 353-372.

[19] B.L. Miller, Finite state continuous time Markov decision processes with a finite planning horizon, SIAM. J. Control Optim, 6 (1968), 266-280.

[20] 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.

[21] S.R. Pliska, Controlled jump processes, Stoch. Process. Appl. 3 (1975), 259-282.

[22] T. Prieto-Rumeau, O. Hernández-Lerma, Selected Topics on Continuous-Time Controlled Markov Chains and Markov Games, Imperial College Press, London, 2012.

[23] T. Prieto-Rumeau, J.M. Lorenzo, Approximating ergodic average reward continuous-time controlled Markov chains, IEEE Trans. Automatic Control, 55 (2010), 201-207.

[24] 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.

[25] J. Shao, K. Zhao, The existence of optimal control for continuous-time Markov decision processes in random environments, arXiv:1906.08423v1 (2019).

[26] J. Shao, The existence of optimal feedback controls for stochastic dynamical systems with regime-switching, arXiv:1906.08425v2, 2019.
[27] D.W. Stroock, S.R.S. Varadhan, Multidimensional diffusion processes, Springer-Verlag, 1979.

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

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