23rd International Symposium on Mathematical Theory of Networks and Systems
Hong Kong University of Science and Technology, Hong Kong, July 16-20, 2018

Hamilton-Jacobi partial differential equations with path-dependent terminal costs under superlinear Lagrangians*

Hidehiro Kaise[1], Takashi Kato[2] and Yusuke Takahashi[3]

Abstract— **We consider variational problems with path-dependent terminal costs. Motivated from Mogulskii's theorem in large deviation theory and dynamic importance sampling for path-dependent rare events, we focus on particular forms of Lagrangians with superlinear growth. By reformulating the variational problem to a value function of a path-dependent deterministic control, we study it by path-dependent dynamic programming methods. Under co-invariant derivative notion on path spaces, the value function is related to a Hamilton-Jacobi partial differential equation (PDE) with a path-dependent terminal condition. Using a viscosity type solution proposed by Lukoyanov for a weak notion, we show that the value function can be characterized as a unique viscosity solution of the Hamilton-Jacobi PDE.**

I. Introduction

Recently there is great attention to problem classes of control theory which cannot be directly treated in a classical framework based on Markovian structures of systems and costs. Systems with delays and pricing of path-dependent options are typical examples. Problems are not limited such ones and the ranges of problem classes are quite large. To investigate non-Markovian systems, mathematical theories of path-space analysis are crucial. Although one might be able to adopt the existing infinite-dimensional analysis theory such as Fréchet derivatives on Banach spaces, it is important for control theory to develop path-space analysis adapted to non-anticipativeness of systems. i-smooth calculus for deterministic systems and functional Itô calculus for stochastic systems are such attempts (see [14] for i-smooth calculus and its applications to functional differential equations and [5] for theory of functional Itô calculus at the early stage). Once theories of non-anticipative path-space analysis are established, one might be able to discuss dynamic programming methods based on those theories. For instance, [14] considers dynamic programming methods for deterministic systems with delays using i-smooth calculus and [10] and [12] develop weak solution notions for hereditary systems. For non-Markovian stochastic systems, the study of dynamic programming by Dupire's derivative and its modifications are quite intense by using techniques of backward stochastic differential equations (cf. [15] and [16] for literature and recent developments).

In this paper, we consider dynamic programming methods on deterministic variational problems with path-dependent terminal costs which typically appear in Mogulskii's theorem in large deviation theory (cf. [4], [6]). Also, such variational problems are asymptotic lower bounds of second moments of unbiased estimators in importance samplings for path-dependent rare events (cf. [7]. See also [2]). The Lagrangians there can have particular forms depending only on a velocity of a trajectory. However they can be superlinearly growing at infinity under mild conditions where Mogulskii's theorem holds. In most of the literature on path-dependent controls, they usually suppose that control spaces and/or running costs are bounded so Lagrangians with superlinear growth cannot be covered. Our goal is to remove the restrictive conditions on running costs and control spaces and lay the groundwork for potential applications to dynamic importance sampling for path-dependent rare events (cf. [7], [8] for dynamic importance sampling for events given by terminal states).

This paper is a part of our project on research of dynamic importance sampling for path-dependent rare events. Using the weak convergence methods of [7], asymptotic optimality of unbiased estimators given by dynamic measure changes is under investigation. Hamilton-Jacobi partial differential equations (PDEs) discussed in this paper may be useful to design optimal dynamic measure changes achieving the asymptotic lower bounds of second moments of unbiased estimators (see [7] and [8] for use of Hamilton-Jacobi and Isaacs PDEs for importance sampling).

The paper is organized as follows. In Section II, we first give a variational problem with a path-dependent cost to be considered. To use a path-dependent dynamic programming method, we reformulate the variational problem to a value function of a path-dependent deterministic control problem. The value function is a function of a pair of a current time and a past history of a state trajectory. We discuss the path-dependent dynamic programming principle and optimality principle. In Section III, we study regularity of value functions and optimal controls which are crucial in the subsequent sections. We show that the value function is Lipschitz continuous. Using optimality principle with the regularity of the value function, we have a bound for optimal controls. In Section IV, we consider a Hamilton-Jacobi PDE, an infinitesimal version of the dynamic programming principle.

*This research was partially supported by JSPS KAKENHI Grant Number 17K05362.

[1]Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan. kaise@sigmath.es.osaka-u.ac.jp

[2]Association of Mathematical Finance Laboratory, 2-10, Kojimachi, Chiyoda, Tokyo 102-0083, Japan. takashi.kato@mathfi-lab.com

[3]National Federation of Workers and Consumers Insurance Cooperatives, 2-12-10, Yoyogi, Shibuya-ku, Tokyo 151-8571 Japan. takahashi_yusuke@zenrosai.coop

Key words. Variational problems, Deterministic optimal control, Dynamic programming methods, Hamilton-Jacobi equations, Viscosity solutions.

2010 Mathematics Subject Classification. 49J15, 49L20, 49L25.

We use co-invariant derivative notion for non-anticipative path-space analysis. By the dynamic programming principle, we formally derive a Hamilton-Jacobi PDE with the path-dependent terminal cost. We introduce verification theorem without proofs. Since the value function is not necessarily differentiable, we need a weak solution notion. In Section V, using the viscosity type weak solution proposed by [12], we characterize the value function as a unique viscosity solution of the Hamilton-Jacobi PDE with the path-dependent terminal cost.

II. Variational problems with path-dependent terminal costs and dynamic programming principle

Let $T > 0$ be a time horizon. Given an initial state $x_0 \in \mathbb{R}^d$, consider the following variational problem:

$$\inf_{\phi \in \mathcal{A}_{x_0}(0,T)} \left\{ \int_0^T L(\dot{\phi}(s))ds + F(\phi) \right\}, \tag{1}$$

where $L : \mathbb{R}^d \to \mathbb{R}$ is a Lagrangian and $F : C([0,T];\mathbb{R}^d) \to \mathbb{R}$ is a path-dependent terminal cost. Here $C([0,T];\mathbb{R}^d)$ is the set of $\mathbb{R}^d$-valued continuous functions on $[0,T]$. We denote by $\mathcal{A}_{x_0}(0,T)$ the set of $\mathbb{R}^d$-valued absolutely continuous functions on $[0,T]$ starting at x_0. The variational problem (1) appears in Mogulskii's theorem in large deviations (cf. [4], [6]). (1) is also the asymptotic lower bound of second moments of unbiased estimators of path-dependent rare events related to Mogulskii's theorem (cf. [7]. See also [2]). It is fundamental in control theory to characterize (1) and find optimal trajectories. For those purposes, we reformulate (1) by a path-dependent deterministic control and use dynamic programming methods.

We prepare some notations which will be needed later. Given $t \in [0,T]$, we set $\mathbf{X}_t = C([0,t];\mathbb{R}^d)$ and denote by x_t an element of $\mathbf{X}_t$ where the subscript of x_t indicates its domain. $x_t(r)$ denotes the value of x_t at $r \in [0,t]$. We understand $\mathbf{X}_0 = \mathbb{R}^d$. Consider $\mathbf{X}_{0,T} = \{(t,x_t);\ t \in [0,T], x_t \in \mathbf{X}_t\}$ and suppose $\mathbf{X}_{0,T}$ is endowed with metric ρ given by $\rho((t,x_t),(s,y_s)) = |t-s| + \max_{0 \le r \le T} |x_t(r \wedge t) - y_s(r \wedge s)|$. Here $a \wedge b = \min\{a,b\}$ for $a,b \in \mathbb{R}$. We suppose that $C([0,T];\mathbb{R}^d)$ is equipped with the sup norm $\|\cdot\|_\infty$.

Given $(t,x_t) \in \mathbf{X}_{0,T}$, consider a controlled ordinary differential equation (ODE) for $\xi : [0,T] \to \mathbb{R}^d$:

$$\begin{aligned} \dot{\xi}(s) &= \beta(s)\ (t \le s \le T), \\ \xi_t &= x_t, \end{aligned} \tag{2}$$

where ξ_t is the restriction of ξ to $[0,t]$. $\beta : [t,T] \to \mathbb{R}^d$ is a control which is a measurable function on $[t,T]$ with $\int_t^T |\beta(s)|ds < \infty$. $\mathcal{B}_{t,T}$ denotes the set of controls on $[t,T]$. Note that (2) is equivalently written by

$$\begin{aligned} \xi(s) &= x_t(t) + \int_t^s \beta(r)dr\ (t \le s \le T), \\ \xi(s) &= x_t(s)\ (0 \le s \le t). \end{aligned}$$

We introduce a criterion $J(t,x_t;\beta)$ given by

$$J(t,x_t;\beta) = \int_t^T L(\beta(s))ds + F(\xi)$$

and then define a value function by

$$V(t,x_t) = \inf_{\beta \in \mathcal{B}_{t,T}} J(t,x_t,\beta). \tag{3}$$

We note that (1) coincides with $V(0,x_0)$.

We always assume that the following conditions hold:

(A1) L is a convex function.

(A2) There exist $p > 1$ and $\nu, C > 0$ such that

$$L(\beta) \ge \nu|\beta|^p - C,\ \forall \beta \in \mathbb{R}^d.$$

(A3) $F : C([0,T];\mathbb{R}^d) \to \mathbb{R}$ is bounded Lipschitz continuous with Lipschitz constant L_F, *i.e.*

$$\|F\|_\infty := \sup_{\phi \in C([0,T];\mathbb{R}^d)} |F(\phi)| < \infty,$$
$$|F(\phi) - F(\psi)| \le L_F \|\phi - \psi\|_\infty,\ \forall \phi,\psi \in C([0,T];\mathbb{R}^d).$$

Remark 2.1: We give comments on (A1) and (A2). Let μ be a probability measure on $\mathbb{R}^d$ satisfying $H(\alpha) := \log \int_{\mathbb{R}^d} e^{\langle \alpha, y \rangle} \mu(dy) < \infty$ for any $\alpha \in \mathbb{R}^d$ where $\langle \cdot, \cdot \rangle$ is Euclidean inner product in $\mathbb{R}^d$. In Mogulskii's theorem, L is given by the Legendre transform of $H(\alpha)$:

$$L(\beta) = \sup_{\alpha \in \mathbb{R}^d} \{\langle \alpha, \beta \rangle - H(\alpha)\},\ \beta \in \mathbb{R}^d.$$

Since L is a supremum of affine functions, L is convex, which corresponds to (A1). We suppose that μ satisfies the following stronger condition:

$$\int_{\mathbb{R}^d} e^{\delta |y|^q} \mu(dy) < \infty \text{ for some } q > 1 \text{ and } \delta > 0. \tag{4}$$

Under (4), it can be seen that L satisfies (A2) for some $p > 1$ and $\nu, C > 0$. Note that (4) holds if the support of μ is compact. For cases of unbounded supports, suppose that μ is absolutely continuous with respect to Lebesgue measure and has density $ce^{p(y)}$ with normalizing constant c. If $p(y) \le -c_1|y|^q + c_2$ $(y \in \mathbb{R}^d)$ for some $q > 1$, $c_1 > 0$ and $c_2 \ge 0$, $\mu(dy) = ce^{p(y)}dy$ satisfies (4). In particular, (4) holds if μ is a normal distribution.

We can have a dynamic programming principle for $V(t,x_t)$. The proof is a straightforward application of arguments of classical deterministic controls. Thus we omit the proof.

Proposition 2.2: For $(t,x_t) \in \mathbf{X}_{0,T}$ and $t \le t+h \le T$, the following holds:

$$V(t,x_t) = \inf_{\beta \in \mathcal{B}_{t,t+h}} \left\{ \int_t^{t+h} L(\beta(s))ds + V(t+h, \xi_{t+h}) \right\}, \tag{5}$$

where $\xi_{t+h} : [0,t+h] \to \mathbb{R}^d$ is the solution of (2).

Remark 2.3: By (A2) and (A3), we can take the infimum in (5) on a smaller class:

$$V(t,x_t)=\inf_{\beta\in\tilde{\mathcal{B}}_{t,t+h}}\left\{\int_t^{t+h}L(\beta(s))ds+V(t+h,\xi_{t+h})\right\}, \tag{6}$$

$$\tilde{\mathcal{B}}_{t,t+h}=\left\{\beta:[t,t+h]\to\mathbb{R}^d;\ \int_t^{t+h}|\beta(s)|^p ds\le M\right\},$$

where M is a constant depending on ν, C, T and $\|F\|_\infty$.

The value function $V(t,x_t)$ is attained at an optimal control.

Proposition 2.4: Let $(t,x_t)\in\mathbf{X}_{0,T}$ be given. Then there exists $\beta^*\in\tilde{\mathcal{B}}_{t,T}$ such that

$$V(t,x_t)=\int_t^T L(\beta^*(s))ds+F(\xi^*),$$

where ξ^* is the solution of (2) with $\beta=\beta^*$.

Proof. By Remark 2.3, we can find an approximating sequence $\{\beta^n\}_{n=1}^\infty\subset\tilde{\mathcal{B}}_{t,T}$ satisfying $J(t,x_t;\beta^n)\to V(t,x_t)$ as $n\to\infty$. Letting ξ^n be the solution of (2) with $\beta=\beta^n$, we see that $\{\xi^n\}_{n=1}^\infty$ is bounded in $H^{1,p}(t,T)$. Here $H^{1,p}(t,T)$ is the Sobolev space of functions $\eta:[t,T]\to\mathbb{R}^d$ with norm $\|\eta\|_{H^{1,p}}=(\int_t^T|\eta(s)|^p ds+\int_t^T|\dot\eta(s)|^p ds)^{1/p}$. Note that $\eta\in H^{1,p}(t,T)$ if and only if η is absolutely continuous on $[t,T]$ and $\int_t^T|\dot\eta(s)|^p ds<\infty$ (cf. [3] for details about $H^{1,p}(t,T)$). By the weak compactness of bounded sets in Banach spaces, ξ^n converges to some $\xi^*\in H^{1,p}(t,T)$ as $n\to\infty$ $H^{1,p}(t,T)$-weakly via a subsequence. We still denote by $\{\xi^n\}_{n=1}^\infty$ the subsequence for simplicity of notations. We extend ξ^* to a function on $[0,T]$ by $\xi^*(r)=x_t(r)$ $(0\le r\le t)$. Set $\beta^*(s)=\dot\xi^*(s)$ $(s\in[t,T])$. By [3, Thm 3.5] with (A1), we can have

$$\int_t^T L(\beta^*(s))ds\le\liminf_{n\to\infty}\int_t^T L(\beta^n(s))ds. \tag{7}$$

Note that $\xi^n(s)$ converges to $\xi^*(s)$ pointwise as $n\to\infty$ by the $H^{1,p}(t,T)$-weak convergence. On the other hand, by using $\beta^n\in\tilde{\mathcal{B}}_{t,T}$, we can show that $\{\xi^n\}_{n=1}^\infty$ is uniformly bounded and equi-continuous on $[t,T]$. By Ascoli-Arzelà theorem, ξ^n converges to ξ^* uniformly on $[t,T]$, thus uniformly on $[0,T]$. By the Lipschitz continuity of F, we have $F(\xi^n)\to F(\xi^*)$ as $n\to\infty$. Combining this with (7), we can have

$$J(t,x_t;\beta^*)\le\liminf_{n\to\infty}J(t,x_t;\beta^n).$$

Hence β^* is an optimal control of $V(t,x_t)$. ■

The following is an extension of the optimality principle along optimal trajectories known in classical deterministic control.

Proposition 2.5: Given $(t,x_t)\in\mathbf{X}_{0,T}$, let $\beta^*\in\mathcal{B}_{t,T}$ be an optimal control of $V(t,x_t)$ and ξ^* be the solution of (2) with $\beta=\beta^*$. For $t\le s\le T$, let ξ_s^* is the restriction of ξ^* to $[0,s]$. Then the following holds: For $s\le s+h\le T$,

$$V(s,\xi_s^*)=\int_s^{s+h}L(\beta^*(s))ds+V(s+h,\xi_{s+h}^*), \tag{8}$$

where ξ_{s+h}^* is the restriction of ξ^* to $[0,s+h]$.

Proof. By Proposition 2.2, we can immediately have

$$V(s,\xi_s^*)\le\int_s^{s+h}L(\beta^*(r))dr+V(s+h,\xi_{s+h}^*). \tag{9}$$

Since β^* is an optimal control for $V(t,x_t)$, we have

$$\begin{aligned}V(t,x_t)&=\int_t^T L(\beta^*(r))dr+F(\xi^*)\\&=\int_t^s L(\beta^*(r))dr+\int_s^T L(\beta^*(r))dr+F(\xi^*).\end{aligned}$$

By Proposition 2.2, we see that

$$V(t,x_t)\le\int_t^s L(\beta^*(r))dr+V(s,\xi_s^*).$$

Thus we can have

$$\int_s^T L(\beta^*(r))dr+F(\xi^*)\le V(s,\xi_s^*),$$

which implies that

$$\begin{aligned}V(s,\xi_s^*)&\ge\int_s^{s+h}L(\beta^*(r))dr+\int_{s+h}^T L(\beta^*(r))dr+F(\xi^*)\\&\ge\int_s^{s+h}L(\beta^*(r))dr+V(s+h,\xi_{s+h}^*).\end{aligned} \tag{10}$$

Note that the second inequality is implied by the definition of $V(s+h,\xi_{t+h}^*)$. Hence, by (9) and (10), we obtain (8). ■

III. Regularity of value functions and optimal controls

We investigate regularity of $V(t,x_t)$ and the optimal controls which will be need to discuss dynamic programming equations for $V(t,x_t)$.

Proposition 3.1: V is Lipschitz continuous in the following sense: There exist $C_1,C_2>0$ such that for any $(t,x_t),(t,y_t)\in\mathbf{X}_{0,T}$ and $t\le t+h\le T$,

$$|V(t,x_t)-V(t,y_t)|\le C_1\|x_t-y_t\|_\infty, \tag{11}$$

$$|V(t+h,x_{t+h}(\cdot\wedge t))-V(t,x_t)|\le C_2h, \tag{12}$$

where $x_{t+h}(r\wedge t):=x_t(r\wedge t)$ $(0\le r\le t+h)$.

Proof. Let $(t,x_t),(t,y_t)\in\mathbf{X}_{0,T}$. We may suppose that $V(t,x_t)\le V(t,y_t)$ without loss of generalities. Letting $\beta^*\in\tilde{\mathcal{B}}_{t,T}$ be an optimal control of $V(t,x_t)$, we have

$$\begin{aligned}V(t,y_t)-V(t,x_t)&\le J(t,y_t;\beta^*)-J(t,x_t;\beta^*)\\&\le F(\eta^*)-F(\xi^*),\end{aligned}$$

where ξ^* (*resp.* η^*) is the solution of (2) with $\xi_t^*=x_t$ (*resp.* $\eta_t^*=y_t$) and $\beta=\beta^*$. Since $\eta^*(s)-\xi^*(s)=y_t(s)-x_t(s)$ $(0\le s\le t)$ and $\eta^*(s)-\xi^*(s)=y_t(t)-x_t(t)$ $(t\le s\le T)$, we have

$$F(\eta^*)-F(\xi^*)\le L_F\|x_t-y_t\|_\infty.$$

Thus we can obtain (11).

Let $(t, x_t) \in \mathbf{X}_{0,T}$ and $t \le t+h \le T$. By (6), we have

$$\begin{aligned} &V(t, x_t) - V(t+h, x_{t+h}(\cdot \wedge t)) \\ &= \inf_{\beta \in \tilde{\mathcal{B}}_{t,t+h}} \left\{ \int_t^{t+h} L(\beta(s)) ds \right. \\ &\qquad \left. + V(t+h, \xi_{t+h}) - V(t+h, x_{t+h}(\cdot \wedge t)) \right\}. \quad (13) \end{aligned}$$

Taking $\beta(s) \equiv 0$ in the right hand side of (13), we have

$$V(t, x_t) - V(t+h, x_{t+h}(\cdot \wedge t)) \le L(0)h \le |L(0)|h.$$

Take an optimal β^* of $V(t, x_t)$. Then we have by Proposition 2.5 and (A2)

$$\begin{aligned} &V(t, x_t) - V(t+h, x_{t+h}(\cdot \wedge t)) \\ &= \int_t^{t+h} L(\beta^*(s)) ds + V(t+h, \xi^*_{t+h}) - V(t+h, x_{t+h}(\cdot \wedge t)) \\ &\ge \nu \int_t^{t+h} |\beta^*(s)|^p ds - Ch \\ &\qquad + V(t+h, \xi^*_{t+h}) - V(t+h, x_{t+h}(\cdot \wedge t)), \end{aligned}$$

where ξ^*_{t+h} is the solution of (2) with $\beta = \beta^*$. By (11), we have

$$\begin{aligned} &|V(t+h, \xi^*_{t+h}) - V(t+h, x_{t+h}(\cdot \wedge t))| \\ &\le C_1 \|\xi^*_{t+h} - x_{t+h}(\cdot \wedge t)\|_\infty \le C_1 \int_t^{t+h} |\beta^*(s)| ds. \end{aligned}$$

Thus we have

$$\begin{aligned} &V(t, x_t) - V(t+h, x_{t+h}(\cdot \wedge t)) \\ &\qquad \ge \int_t^{t+h} \{\nu |\beta^*(s)|^p - C_1 |\beta^*(s)|\} ds - Ch. \end{aligned}$$

Since $|\beta|^p - C_1|\beta| \to \infty$ as $|\beta| \to \infty$, there exists $\hat{C} > 0$ such that $|\beta|^p - C_1|\beta| \ge -\hat{C}$ ($\beta \in \mathbb{R}^d$). Thus we can have

$$V(t, x_t) - V(t+h, x_{t+h}(\cdot \wedge t)) \ge -\hat{C}h - Ch. \blacksquare$$

The following estimate for optimal controls is crucial for the later arguments.

Proposition 3.2: Let $(t, x_t) \in \mathbf{X}_{0,T}$ and $\beta^* \in \mathcal{B}_{t,T}$ be an optimal control of $V(t, x_t)$. Then there exists a constant K depending on ν, C, p, C_1 and C_2 such that

$$|\beta^*(s)| \le K, \ \ a.e. \, s \in [t, T].$$

Proof. Let $t \le s \le s+h \le T$. By Proposition 2.5, we have

$$V(s, \xi^*_s) = \int_s^{s+h} L(\beta^*(s)) ds + V(s+h, \xi^*_{s+h}),$$

where ξ^* is the solution of (2) with $\beta = \beta^*$ and ξ^*_{s+h} is the restriction of ξ^* to $[0, s+h]$. This equation can be written by

$$\int_s^{s+h} L(\beta^*(s)) ds = V(s, \xi^*_s) - V(s+h, \xi^*_{s+h}).$$

Note that

$$\begin{aligned} &|V(s+h, \xi^*_{s+h}) - V(s, \xi^*_s)| \\ &\le |V(s+h, \xi^*_{s+h}) - V(s+h, \xi^*_{s+h}(\cdot \wedge s)| \\ &\quad + |V(s+h, \xi^*_{s+h}(\cdot \wedge s) - V(s, \xi^*_s)|, \end{aligned}$$

where $\xi^*_{s+h}(r \wedge s) = \xi^*_s(r \wedge s)$ $(0 \le r \le s+h)$. By Proposition 3.1, we have

$$\begin{aligned} &|V(s+h, \xi^*_{s+h}) - V(s+h, \xi^*_{s+h}(\cdot \wedge s)| \\ &\quad + |V(s+h, \xi^*_{s+h}(\cdot \wedge s) - V(s, \xi^*_s)| \\ &\le C_1 \int_s^{s+h} |\beta^*(r)| dr + C_2 h. \end{aligned}$$

Recalling (A2), we have

$$\int_s^{s+h} L(\beta^*(r)) dr \ge \nu \int_s^{s+h} |\beta^*(r)|^p - Ch.$$

Thus we obtain

$$\nu \int_s^{s+h} |\beta^*(r)|^p dr - Ch \le C_1 \int_s^{s+h} |\beta^*(r)| dr + C_2 h.$$

Dividing the above inequality by h and sending $h \to 0+$, we have

$$\nu |\beta^*(s)|^p - C_1 |\beta^*(s)| \le C + C_2.$$

Since $\nu|\beta|^p - C_1|\beta| \to \infty$ $(|\beta| \to \infty)$, there exist K depending on ν, p, C_1, C_2 and C such that

$$|\beta^*(s)| \le K. \blacksquare$$

Remark 3.3: By Proposition 2.5 with Proposition 3.2, (5) can be reduced to

$$V(t, x_t) = \inf_{\beta \in \mathcal{B}^K_{t,t+h}} \left\{ \int_t^{t+h} L(\beta(s)) ds + V(t+h, \xi_{t+h}) \right\}, \quad (14)$$

where

$$\mathcal{B}^K_{t,t+h} = \{\beta \in \mathcal{B}_{t,t+h};\ |\beta(s)| \le K,\ a.e. \, s \in [t, t+h]\}.$$

IV. Hamilton-Jacobi PDEs with path-dependent terminal costs

To derive a dynamic programming PDE from (5), we use co-invariant derivatives for a notion of derivatives on path spaces (cf. [14]). To introduce the derivative notion, we need some function spaces. Let $Lip(t, x_t)$ be the set of continuous functions $y : [0, T] \to \mathbb{R}^d$ such that y is Lipschitz continuous on $[t, T]$ and $y_t = x_t$ where y_t is the restriction of y to $[0, t]$. For $\kappa > 0$, let $Lip^{(\kappa)}(t, x_t)$ be the set of $y \in Lip(t, x_t)$ whose Lipschitz constant on $[t, T]$ is κ.

Definition 4.1 (cf. [14]): Let $\varphi : \mathbf{X}_{0,T} \to \mathbb{R}$ and $(t, x_t) \in \mathbf{X}_{0,T-} := \{(t, x_t) \in \mathbf{X}_{0,T};\ 0 \le t < T\}$. φ is *co-invariant differentiable at* (t, x_t) if there exist $a \in \mathbb{R}$ and $p \in \mathbb{R}^d$ such that for any $y \in Lip(t, x_t)$,

$$\begin{aligned} \varphi(t+h, y_{t+h}) = \varphi(t, x_t) + ah + \langle p, y(t+h) - x_t(t) \rangle \\ + h\omega(h; (t, x_t), y), \quad (15) \end{aligned}$$

where $\omega(\cdot; (t, x_t), y) : (0, T-t] \to \mathbb{R}$ is a continuous function depending on (t, x_t) and y such that $\omega(h; (t, x_t), y) \to$

0 $(h \to 0)$. a and p are called *co-invariant derivatives* and denoted by $\partial_t \varphi(t,x_t)$ and $\nabla_{x_t}\varphi(t,x_t)$, respectively.

$\varphi : \mathbf{X}_{0,T} \to \mathbb{R}$ is in $\mathcal{C}^1(\mathbf{X}_{0,T})$ if φ is co-invariant differentiable on $\mathbf{X}_{0,T-}$, and φ, $\partial_t\varphi$ and $\nabla_{x_t}\varphi$ are continuous. For the later arguments, we define $\mathcal{C}^1_u(\mathbf{X}_{0,T})$ by the set of $\varphi \in \mathcal{C}^1(\mathbf{X}_{0,T})$ satisfying for each $(t,x_t) \in \mathbf{X}_{0,T-}$ and $\kappa > 0$,

$$\sup_{y \in Lip^{(\kappa)}(t,x_t)} |\omega(h;(t,x_t),y)| \to 0 \ (h \to 0+),$$

where $\omega(h,(t,x_t),y)$ is from (15).

Remark 4.2: Let $\psi : [0,T]\times\mathbb{R}^d \to \mathbb{R}$ be a C^1-function in the conventional sense with partial derivatives $(\partial\psi/\partial t)(t,x)$ and $D_x\psi(t,x) = ((\partial\psi/\partial x_1)(t,x),\dots,(\partial\psi/\partial x_d)(t,x))$ $((t,x) = (t,x_1,\dots,x_d) \in [0,T]\times\mathbb{R}^d)$. Setting $\varphi(t,x_t) = \psi(t,x_t(t))$ $((t,x_t) \in \mathbf{X}_{0,T})$, it is not difficult to see that $\varphi \in \mathcal{C}^1_u(\mathbf{X}_{0,T})$ and $\partial_t\varphi(t,x_t) = (\partial\psi/\partial t)(t,x_t(t))$ and $\nabla_{x_t}\varphi(t,x_t) = D_x\psi(t,x_t(t))$.

Using (5) with the co-invariant derivative notion, we can formally derive a Hamilton-Jacobi PDE with a path-dependent terminal cost:

$$\partial_t V(t,x_t) + \inf_{\beta\in\mathbb{R}^d}\{\langle\beta,\nabla_{x_t}V(t,x_t)\rangle + L(\beta)\} = 0, \qquad (t,x_t)\in\mathbf{X}_{0,T-}, \tag{16}$$

$$V(T,x_T) = F(x_T),\ x_T \in \mathbf{X}_T. \tag{17}$$

A verification theorem holds under the existence of $\mathcal{C}^1(\mathbf{X}_{0,1})$-solutions. The proof is similar to that of classical deterministic control (cf. [11]).

Theorem 4.3: Suppose $w : \mathbf{X}_{0,T} \to \mathbb{R}$ is in $\mathcal{C}^1(\mathbf{X}_{0,T})$ and satisfies (16) and (17). Then the followings hold:
(i) $w(t,x_t) \le J(t,x_t;\beta)$ for any $\beta \in \mathcal{B}_{t,T}$.
(ii) Let $\beta^*(s,y_s) \in \arg\min_{\beta\in\mathbb{R}^d}\{\langle\beta,\nabla_{x_t}w(s,y_s)\rangle + L(\beta)\}$ for $(s,y_s)\in\mathbf{X}_{0,T-}$. Suppose (2) has a solution $\xi = \xi^*$ with the choice $\beta(s) = \beta^*(s) = \beta^*(s,\xi^*_s)$ and ξ^* is Lipschitz on $[t,T]$. Then $w(t,x_t) = J(t,x;\beta^*)$.

V. Viscosity characterization of value functions

Since the value function given by (3) is not necessarily co-invariant differentiable, (16) has to be understood by a weak solution. Here we adopt a viscosity solution notion proposed in [12] and extend it to unbounded control spaces.

To define a path-dependent viscosity notion following the idea of [12], we need function spaces given below. Let $a \ge 1$ be given. Let D_k $(k = 1,2,\dots)$ and D be defined by

$$\begin{aligned} D_k =& \{(t,x_t)\in\mathbf{X}_{0,T};\ x_t \text{ is absolutely continuous on } [0,t],\\ & |x_t(0)| \le k, |\dot{x}_t(r)| \le ka(1+\|x_r\|_\infty)\ \textit{a.e.}\ r\in[0,t]\},\\ D =& \bigcup_{k=1}^{\infty} D_k. \end{aligned}$$

Remark 5.1: Note that closed balls in $\mathbf{X}_{0,T}$ are not compact. On the other hand, it can be seen that D_k is compact in $\mathbf{X}_{0,T}$ and D is dense in $\mathbf{X}_{0,T}$. D_k is used for local maximum/minimum arguments in viscosity theory of [12] (cf. [1] for the conventional viscosity theory).

Definition 5.2 (cf.[12]): Suppose that $w : \mathbf{X}_{0,T} \to \mathbb{R}$ satisfies (11) and (12). w is a *viscosity subsolution* of (16) if the following condition holds: Let $\varphi \in \mathcal{C}^1_u(\mathbf{X}_{0,T})$ and $k\in\mathbb{N}$. If $(\hat{t},\hat{x}_{\hat{t}}) \in \mathbf{X}_{0,T-}$ is a maximum point of $w-\varphi$ on D_k and

$$|\nabla_{x_t}\varphi(\hat{t},\hat{x}_{\hat{t}})| \le C_1 + 2C_2, \tag{18}$$

then

$$\partial_t\varphi(\hat{t},\hat{x}_{\hat{t}}) + \inf_{\beta\in\mathbb{R}^d}\{\langle\beta,\nabla_{x_t}\varphi(\hat{t},\hat{x}_{\hat{t}})\rangle + L(\beta)\} \ge 0. \tag{19}$$

w is a *viscosity supersolution* if the above condition holds when replacing "maximum" with "minimum" and "$\ge$" with "$\le$" in (19). w is a *viscosity solution* if w is a viscosity sub and supersolution.

Remark 5.3: In the conventional viscosity theory, inheritances of Lipschitz constants of value functions to test functions are useful to treat unbounded control spaces (cf. [9, Chap. VII]). However D_k is too small to obtain such inheritances for general test functions. We note that a particular test function is used in the proof of our comparison theorem and we will see that it inherits the Lipschitz constant in the sense of (18) (see (30) in the proof of Theorem 5.6). Thus it suffices to consider a class of test functions with (18) which makes the proof of the existence of viscosity solutions easier (see the proof of Theorem 5.5).

Classical subsolutions and supersolutions under co-invariant derivative notions are viscosity subsolutions and supersolutions, respectively.

Proposition 5.4: Let $w \in \mathcal{C}^1(\mathbf{X}_{0,T})$ be a classical subsolution (*resp.* supersolution) of (16) satisfying (11) and (12). If we take a in the definition of D sufficiently large, then w is a viscosity subsolution (*resp.* supersolution) of (16).

Proof. Since the proof similar to the current case was already given in [13], we give an outline of the proof. We only prove the subsolution part since the proof for supersolutions can be obtained in a way similar to that of subsolutions.

Let $w \in \mathcal{C}^1(\mathbf{X}_{0,T})$ be a subsolution of (16), *i.e.*

$$\partial_t w(t,x_t) + G(\nabla_{x_t}w(t,x_t)) \ge 0,\ (t,x_t)\in\mathbf{X}_{0,T-},$$

where $G(p) = \inf_{\beta\in\mathbb{R}^d}\{\langle\beta,p\rangle + L(\beta)\}$. Let $\varphi\in\mathcal{C}^1_u(\mathbf{X}_{0,T})$ and $k\in\mathbb{N}$. Let $(\hat{t},\hat{x}_{\hat{t}})$ be a maximum point of $w-\varphi$ on D_k and $|\nabla_{x_t}\varphi(\hat{t},\hat{x}_{\hat{t}})| \le C_1 + 2C_2$. We note that

$$\partial_t w(\hat{t},\hat{x}_{\hat{t}}) + G(\nabla_{x_t}w(\hat{t},\hat{x}_{\hat{t}})) \ge 0. \tag{20}$$

We consider the case where $\nabla_{x_t}w(\hat{t},\hat{x}_{\hat{t}}) = \nabla_{x_t}\varphi(\hat{t},\hat{x}_{\hat{t}})$ Noting that

$$w(t,x_t) - w(\hat{t},\hat{x}_{\hat{t}}) \le \varphi(t,x_t) - \varphi(\hat{t},\hat{x}_{\hat{t}}),\ (t,x_t)\in D_k \tag{21}$$

and taking $y = \hat{x}_{\hat{t}}(\cdot\wedge\hat{t}) \in Lip(\hat{t},\hat{x}_{\hat{t}})$ for the co-invariant differentiability, we can have

$$\partial_t w(\hat{t},\hat{x}_{\hat{t}}) \le \partial_t\varphi(\hat{t},\hat{x}_{\hat{t}}).$$

Thus, by (20), we obtain

$$\partial_t\varphi(\hat{t},\hat{x}_{\hat{t}}) + G(\nabla_{x_t}\varphi(\hat{t},\hat{x}_{\hat{t}})) \ge 0.$$

We next consider the case where $\nabla_{x_t} w(\hat{t}, \hat{x}_{\hat{t}}) \neq \nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}})$. Set $p_w = \nabla_{x_t} w(\hat{t}, \hat{x}_{\hat{t}})$ and $p_\varphi = \nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}})$ for simplifications of notations. We define $z \in Lip(\hat{t}, \hat{x}_{\hat{t}})$ by

$$z(r) = \begin{cases} \hat{x}_{\hat{t}}(r) & (0 \le r \le \hat{t}), \\ \hat{x}_{\hat{t}}(\hat{t}) + \alpha(r - \hat{t}) & (\hat{t} \le r \le T), \end{cases}$$

where

$$\alpha = \frac{G(p_w) - G(p_\varphi)}{|p_w - p_\varphi|^2}(p_w - p_\varphi).$$

Using (11) and (12) for the co-invariant differentiability of w at $(\hat{t}, \hat{x}_{\hat{t}})$ with the choice $y \in Lip(\hat{t}, \hat{x}_{\hat{t}})$ where y is any straight line on $[\hat{t}, T]$, we can have $|p_w| \le C_1$. Using this estimate of $|p_w|$ and $|p_\varphi| \le C_1 + 2C_2$ and (A2), we can see that there exist a constant M depending on C_1, C_2, ν, p, C such that

$$|\alpha| = \frac{|G(p_w) - G(p_\varphi)|}{|p_w - p_\varphi|} \le M$$

Thus, if we take $a \ge M$, we can see $(\hat{t} + h, z_{\hat{t}+h}) \in D_k$. Then, using (21) with co-invariant differentiability of w and φ, we have

$$\begin{aligned} &\partial_t w(\hat{t}, \hat{x}_{\hat{t}})h + \langle p_w, z(\hat{t} + h) - \hat{x}_{\hat{t}}(\hat{t})\rangle + o(h) \\ &\le \partial_t \varphi(\hat{t}, \hat{x}_{\hat{t}})h + \langle p_\varphi, z(\hat{t} + h) - \hat{x}_{\hat{t}}(\hat{t})\rangle + o(h), \end{aligned}$$

which implies

$$\partial_t w(\hat{t}, \hat{x}_{\hat{t}}) + \langle p_w, \alpha\rangle \le \partial_t\varphi(\hat{t}, \hat{x}_{\hat{t}}) + \langle p_\varphi, \alpha\rangle.$$

Recalling (20) and using the above inequality, we can have

$$\begin{aligned} 0 &\le \partial_t w(\hat{t}, \hat{x}_{\hat{t}}) + G(p_w) \\ &\le \partial_t\varphi(\hat{t}, \hat{x}_{\hat{t}}) + \langle p_\varphi - p_w, \alpha\rangle + G(p_w) \\ &= \partial_t\varphi(\hat{t}, \hat{x}_{\hat{t}}) + G(p_\varphi). \ \blacksquare \end{aligned}$$

The value function can be characterized as a unique viscosity solution of the Hamilton-Jacobi PDE with the path-dependent terminal cost.

Theorem 5.5: Suppose (A1)–(A3) hold. Take a in the definition of D sufficiently large. Then $V(t, x_t)$ is a unique viscosity solution of (16) with (17) satisfying (11) and (12).

Proof. We will show that $V(t, x_t)$ is a viscosity subsolution of (16). Let $\varphi \in \mathcal{C}^1_u(\mathbf{X}_{0,T})$ and $k \in \mathbb{N}$. Let $(\hat{t}, \hat{x}_{\hat{t}}) \in \mathbf{X}_{0,T-}$ be a maximum point of $V - \varphi$ on D_k and suppose $|\nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}})| \le C_1 + 2C_2$. Take $\tilde{\beta} \in \arg\min_{\beta\in\mathbb{R}^d}\{\langle\beta, \nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}})\rangle + L(\beta)\}$. By (A2), note that there exists $M > 0$ depending on $\nu, C, L(0)$ such that

$$|\tilde{\beta}|^p \le M(1 + |\nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}})|^q), \tag{22}$$

where $q > 1$ is the conjugate of p, *i.e.* $1/p + 1/q = 1$. Since $|\nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}})| \le C_1 + 2C_2$, there exists $\bar{M} = \bar{M}(p, M, C_1, C_2)$ such that

$$|\tilde{\beta}| \le \bar{M}.$$

If we take $\beta(s) \equiv \tilde{\beta}$ at (5) with $(t, x_t) = (\hat{t}, \hat{x}_{\hat{t}})$, we have

$$V(\hat{t}, \hat{x}_{\hat{t}}) \le L(\tilde{\beta})h + V(\hat{t} + h, \tilde{\xi}_{\hat{t}+h}),$$

where $\tilde{\xi}$ is the solution of (2) with $(t, x_t) = (\hat{t}, \hat{x}_{\hat{t}})$ and $\beta(s) \equiv \tilde{\beta}$. This can be

$$0 \le L(\tilde{\beta})h + V(\hat{t} + h, \tilde{\xi}_{\hat{t}+h}) - V(\hat{t}, \hat{x}_{\hat{t}}). \tag{23}$$

Take a such that $a \ge \bar{M}$. Then we can see that $(\hat{t}+h, \tilde{\xi}_{\hat{t}+h}) \in D_k$. Recalling $(\hat{t}, \hat{x}_{\hat{t}})$ is a maximum point of $V - \varphi$ on D_k, we have from (23)

$$\begin{aligned} 0 &\le L(\tilde{\beta})h + V(\hat{t} + h, \tilde{\xi}_{\hat{t}+h}) - V(\hat{t}, \hat{x}_{\hat{t}}) \\ &\le L(\tilde{\beta})h + \varphi(\hat{t} + h, \tilde{\xi}_{\hat{t}+h}) - \varphi(\hat{t}, \hat{x}_{\hat{t}}) \\ &= L(\tilde{\beta})h + \partial_t\varphi(\hat{t}, \hat{x}_{\hat{t}})h + \langle\nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}}), \tilde{\beta}\rangle h \\ &\qquad + h\omega(h; (\hat{t}, \hat{x}_{\hat{t}}), \tilde{\xi}). \end{aligned}$$

Dividing the above inequality by h and sending $h \to 0+$, we have

$$0 \le \partial_t\varphi(\hat{t}, \hat{x}_{\hat{t}}) + \langle\nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}}), \tilde{\beta}\rangle + L(\tilde{\beta}).$$

Since $\tilde{\beta} \in \arg\min_{\beta\in\mathbb{R}^d}\{\langle\nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}}), \beta\rangle + L(\beta)\}$, we can have

$$0 \le \partial_t\varphi(\hat{t}, \hat{x}_{\hat{t}}) + \inf_{\beta\in\mathbb{R}^d}\{\langle\nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}}), \beta\rangle + L(\beta)\}.$$

Thus $V(t, x_t)$ is a viscosity subsolution of (16).

We next show that $V(t, x_t)$ is a viscosity supersolution of (16). Let $\varphi \in \mathcal{C}^1_u(\mathbf{X}_{0,T})$ and $k \in \mathbb{N}$. Let $(\hat{t}, \hat{x}_{\hat{t}}) \in \mathbf{X}_{0,T-}$ be a minimum point of $V - \varphi$ on D_k and suppose $|\nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}})| \le C_1 + 2C_2$. By (14), we note that the following holds;

$$0 = \inf_{\beta\in\mathcal{B}^K_{\hat{t},\hat{t}+h}}\left\{\int_{\hat{t}}^{\hat{t}+h} L(\beta(s))ds + V(\hat{t} + h, \hat{\xi}^\beta_{\hat{t}+h}) - V(\hat{t}, \hat{x}_{\hat{t}})\right\}, \tag{24}$$

where $\hat{\xi}^\beta_{\hat{t}+h}$ is the solution of (2) with $(t, x_t) = (\hat{t}, \hat{x}_{\hat{t}})$ and $\beta \in \mathcal{B}^K_{\hat{t},\hat{t}+h}$. Taking a such that $a \ge K$, we can see that $(\hat{t}+h, \hat{\xi}^\beta_{\hat{t}+h}) \in D_k$ for any $\beta \in \mathcal{B}^K_{\hat{t},\hat{t}+h}$. Thus, since $(\hat{t}, \hat{x}_{\hat{t}})$ is a minimum point of $V - \varphi$ on D_k, we can have from (24)

$$0 \ge \inf_{\beta\in\mathcal{B}^K_{\hat{t},\hat{t}+h}}\left\{\int_{\hat{t}}^{\hat{t}+h} L(\beta(s))ds + \varphi(\hat{t} + h, \hat{\xi}^\beta_{\hat{t}+h}) - \varphi(\hat{t}, \hat{x}_{\hat{t}})\right\}. \tag{25}$$

Noting that $\hat{\eta}^\beta(r) := \hat{\xi}^\beta_{\hat{t}+h}(r \wedge (\hat{t} + h))$ $(r \in [0, T])$ is Lipschitz continuous on $[\hat{t}, T]$ with Lipschitz constant K because $\|\beta\|_{L^\infty(\hat{t},\hat{t}+h)} \le K$, $\hat{\eta}^\beta \in Lip^{(K)}(\hat{t}, \hat{x}_{\hat{t}})$. Then using $\varphi \in \mathcal{C}^1_u(\mathbf{X}_{0,T})$ with (25), we can have

$$\begin{aligned} 0 &\ge \inf_{\beta\in\mathcal{B}^K_{\hat{t},\hat{t}+h}}\left\{\int_{\hat{t}}^{\hat{t}+h} L(\beta(s))ds + \partial_t\varphi(\hat{t}, \hat{x}_{\hat{t}})h\right. \\ &\qquad \left. + \left\langle\nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}}), \int_{\hat{t}}^{\hat{t}+h}\beta(s)ds\right\rangle + h\omega(h; (\hat{t}, \hat{x}_{\hat{t}}), \hat{\eta}^\beta)\right\} \\ &\ge \inf_{\beta\in\mathcal{B}^K_{\hat{t},\hat{t}+h}}\left\{\int_{\hat{t}}^{\hat{t}+h}\Big[\partial_t\varphi(\hat{t}, \hat{x}_{\hat{t}}) + \langle\nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}}), \beta(s)\rangle\right. \\ &\qquad \left. + L(\beta(s))\Big]ds\right\} - h\sup_{y\in Lip^{(K)}(\hat{t},\hat{x}_{\hat{t}})}|\omega(h; (\hat{t}, \hat{x}_{\hat{t}}), y)| \\ &\ge h\Big[\partial_t\varphi(\hat{t}, \hat{x}_{\hat{t}}) + \inf_{\beta\in\mathbb{R}^d}\{\langle\nabla_{x_t}\varphi(\hat{t}, \hat{x}_{\hat{t}}), \beta\rangle + L(\beta)\}\Big] \\ &\qquad - h\sup_{y\in Lip^{(K)}(\hat{t},\hat{x}_{\hat{t}})}|\omega(h; (\hat{t}, \hat{x}_{\hat{t}}), y)|. \end{aligned}$$

Dividing the above by h and $h \to 0+$, we obtain

$$0 \geq \partial_t \varphi(\hat{t}, \hat{x}_{\hat{t}}) + \inf_{\beta \in \mathbb{R}^d} \{\langle \nabla_{x_t} \varphi(\hat{t}, \hat{x}_{\hat{t}}), \beta \rangle + L(\beta)\}.$$

Hence $V(t, x_t)$ is a viscosity supersolution of (16).

The uniqueness of viscosity solutions is an implication of the comparison theorem given below. ■

Theorem 5.6: Suppose (A1)–(A3) hold. Let v and w be a viscosity subsolution and supersolution of (16) satisfying (11) and (12), respectively. If $v(T, x_T) \leq w(T, x_T)$ for $x_T \in \mathbf{X}_T$, then $v(t, x_t) \leq w(t, x_t)$ for $(t, x_t) \in \mathbf{X}_{0,T}$.

Proof. We follow the arguments of the proof of [12, Theorem 2] and modify them for cases of unbounded control spaces and superlinear Lagrangians. Although we already proved a comparison theorem in [13] with unbounded control spaces (but under a viscosity notion slightly different from Definition 5.2) and here we use the arguments similar to that in [13], we will give a quick proof for the readers' convenience.

Since v and w are continuous and $D = \cup_{k=1}^{\infty} D_k$ is dense in $\mathbf{X}_{0,T}$, it suffices to show that $v \leq w$ on D. On the contrary, suppose that there exists $k \in \mathbb{N}$ such that

$$\delta := \max_{(t,x_t) \in D_k} \{v(t, x_t) - w(t, x_t)\} > 0. \tag{26}$$

Let $0 < \alpha < \delta/4T$. Given $0 < \epsilon < 1$, we define $\Phi_\epsilon : \mathbf{X}_{0,T} \times \mathbf{X}_{0,T} \to \mathbb{R}$ by

$$\Phi_\epsilon((t, x_t), (s, y_s)) = v(t, x_t) - w(s, y_s) - \alpha(2T - t - s) - \frac{1}{2\epsilon} \nu((t, x_t), (s, y_s)),$$

where

$$\nu((t, x_t), (s, y_s)) = |t - s|^2 + |x_t(t) - y_s(s)|^2 + \int_0^T |x_t(r \wedge t) - y_s(r \wedge s)|^2 dr.$$

Let $((t_\epsilon, x^\epsilon_{t_\epsilon}), (s_\epsilon, y^\epsilon_{s_\epsilon}))$ be a maximum point of Φ_ϵ on $D_k \times D_k$. By standard arguments as in comparison theorems of conventional viscosity theory, we can have $(1/2\epsilon)\nu((t_\epsilon, x^\epsilon_{t_\epsilon}), (s_\epsilon, y^\epsilon_{s_\epsilon})) \leq 2(\max_{D_k} |v| + \max_{D_k} |w|)$. Also, by the arguments used in comparison theorems with uniform continuity of v and w on D_k, we can have

$$\rho((t_\epsilon, x^\epsilon_{t_\epsilon}), (s_\epsilon, y^\epsilon_{s_\epsilon})) \to 0 \ (\epsilon \to 0), \tag{27}$$

$$\frac{1}{2\epsilon} \nu((t_\epsilon, x^\epsilon_{t_\epsilon}), (s_\epsilon, y^\epsilon_{s_\epsilon})) \to 0 \ (\epsilon \to 0). \tag{28}$$

We note that $t_\epsilon, s_\epsilon < T$ for sufficiently small ϵ because of (26), (27) and $v(T, x_T) \leq w(T, x_T)$.

Since $(t_\epsilon, x^\epsilon_{t_\epsilon})$ is a maximum point of $\Phi_\epsilon(\cdot, (s_\epsilon, y^\epsilon_{s_\epsilon}))$ on D_k, we have

$$v(t, x_t) - v(t_\epsilon, x^\epsilon_{t_\epsilon}) \leq \phi(t, x_t) - \phi(t_\epsilon, x^\epsilon_{t_\epsilon}), \ (t, x_t) \in D_k, \tag{29}$$

where

$$\phi(t, x_t) = w(s_\epsilon, y^\epsilon_{s_\epsilon}) + \alpha(2T - t - s_\epsilon) + \frac{1}{2\epsilon} \nu((t, x_t), (s_\epsilon, y^\epsilon_{s_\epsilon})).$$

We note that $\phi \in \mathcal{C}^1_u(\mathbf{X}_{0,T})$ and

$$\partial_t \phi(t_\epsilon, x^\epsilon_{t_\epsilon}) = -\alpha + \frac{t_\epsilon - s_\epsilon}{\epsilon},$$

$$\nabla_{x_t} \phi(t_\epsilon, x^\epsilon_{t_\epsilon}) = \frac{1}{\epsilon}(x^\epsilon_{t_\epsilon}(t_\epsilon) - y^\epsilon_{s_\epsilon}(s_\epsilon)) + \frac{1}{\epsilon} \int_{t_\epsilon}^T \{x^\epsilon_{t_\epsilon}(t_\epsilon) - y^\epsilon_{s_\epsilon}(r \wedge s_\epsilon)\} dr.$$

Similarly, since $(s_\epsilon, y^\epsilon_{s_\epsilon})$ is a minimum point of $-\Phi_\epsilon((t_\epsilon, x^\epsilon_{t_\epsilon}), \cdot)$ on D_k, we have

$$\psi(s, y_s) - \psi(s_\epsilon, y^\epsilon_{s_\epsilon}) \leq w(s, y_s) - w(s_\epsilon, y^\epsilon_{s_\epsilon}), \ (s, y_s) \in D_k,$$

where

$$\psi(s, y_s) = v(t_\epsilon, x^\epsilon_{t_\epsilon}) - \alpha(2T - t_\epsilon - s) - \frac{1}{2\epsilon} \nu((t_\epsilon, x^\epsilon_{t_\epsilon}), (s, y_s)).$$

Noting $\psi \in \mathcal{C}^1_u(\mathbf{X}_{0,T})$, we can have

$$\partial_t \psi(s_\epsilon, y^\epsilon_{s_\epsilon}) = \alpha + \frac{t_\epsilon - s_\epsilon}{\epsilon},$$

$$\nabla_{x_t} \psi(s_\epsilon, y^\epsilon_{s_\epsilon}) = \frac{1}{\epsilon}(x^\epsilon_{t_\epsilon}(t_\epsilon) - y^\epsilon_{s_\epsilon}(s_\epsilon)) + \frac{1}{\epsilon} \int_{s_\epsilon}^T \{x^\epsilon_{t_\epsilon}(r \wedge t_\epsilon) - y^\epsilon_{s_\epsilon}(s_\epsilon)\} dr.$$

Set $p_\epsilon = \nabla_{x_t} \phi(t_\epsilon, x^\epsilon_{t_\epsilon})$ and $q_\epsilon = \nabla_{x_t} \psi(s_\epsilon, y^\epsilon_{s_\epsilon})$. We will show that

$$|p_\epsilon| \leq C_1 + 2C_2, \ |q_\epsilon| \leq C_1 + 2C_2. \tag{30}$$

where C_1 and C_2 are from (11) and (12) which v and w satisfy. By $\phi \in \mathcal{C}^1(\mathbf{X}_{0,T})$ and (29), we have

$$\begin{aligned} &v(t_\epsilon + h, \zeta_{t_\epsilon + h}) - v(t_\epsilon, x^\epsilon_{t_\epsilon}) \\ &\leq \partial_t \phi(t_\epsilon, x^\epsilon_{t_\epsilon}) h + \langle \nabla_{x_t} \phi(t_\epsilon, x^\epsilon_{t_\epsilon}), \zeta(t_\epsilon + h) - x^\epsilon_{t_\epsilon}(t_\epsilon) \rangle \\ &\quad + h\omega(h; (t_\epsilon, x^\epsilon_{t_\epsilon}), \zeta), \\ &\forall \zeta \in Lip(t_\epsilon, x^\epsilon_{t_\epsilon}) \text{ with } (t_\epsilon + h, \zeta_{t_\epsilon + h}) \in D_k. \end{aligned}$$

Noting that v satisfies (11) and (12), we can have

$$\begin{aligned} &- C_1 \|\zeta_{t_\epsilon + h} - x^\epsilon_{t_\epsilon + h}(\cdot \wedge t_\epsilon)\|_\infty - C_2 h \\ &\leq \partial_t \phi(t_\epsilon, x^\epsilon_{t_\epsilon}) h + \langle \nabla_{x_t} \phi(t_\epsilon, x^\epsilon_{t_\epsilon}), \zeta(t_\epsilon + h) - x^\epsilon_{t_\epsilon}(t_\epsilon) \rangle \\ &\quad + h\omega(h; (t_\epsilon, x^\epsilon_{t_\epsilon}), \zeta), \\ &\quad \forall \zeta \in Lip(t_\epsilon, x^\epsilon_{t_\epsilon}) \text{ with } (t_\epsilon + h, \zeta_{t_\epsilon + h}) \in D_k. \end{aligned} \tag{31}$$

By arguments similar to the above, we also can obtain

$$\begin{aligned} &\partial_t \psi(s_\epsilon, y^\epsilon_{s_\epsilon}) h + \langle \nabla_{x_t} \psi(s_\epsilon, y^\epsilon_{s_\epsilon}), \zeta(s_\epsilon + h) - y^\epsilon(s_\epsilon) \rangle \\ &+ h\omega(h; (s_\epsilon, y^\epsilon_{s_\epsilon}), \zeta) \\ &\leq C_1 \|\zeta_{s_\epsilon + h} - y^\epsilon_{s_\epsilon + h}(\cdot \wedge s_\epsilon)\|_\infty + C_2 h, \\ &\quad \forall \zeta \in Lip(s_\epsilon, y^\epsilon_{s_\epsilon}) \text{ with } (s_\epsilon + h, \zeta_{s_\epsilon + h}) \in D_k. \end{aligned} \tag{32}$$

Let $\zeta \in Lip(t_\epsilon, x^\epsilon_{t_\epsilon})$ be given by

$$\zeta(r) = \begin{cases} x^\epsilon_{t_\epsilon}(r) & (0 \leq r \leq t_\epsilon), \\ x^\epsilon_{t_\epsilon}(t_\epsilon) & (t_\epsilon \leq r \leq T). \end{cases}$$

Since $(t_\epsilon + h, \zeta_{t_\epsilon + h}) \in D_k$, we have from (31)

$$-C_2 h \leq \partial_t \phi(t_\epsilon, x^\epsilon_{t_\epsilon}) h + h\omega(h; (t_\epsilon, x^\epsilon_{t_\epsilon}), \zeta),$$

which implies

$$-C_2 + \alpha \leq \frac{t_\epsilon - s_\epsilon}{\epsilon}.$$

Using (32) with an argument similar to the above, we can have

$$\partial_t \psi(s_\epsilon, y^\epsilon_{s_\epsilon}) \leq C_2,$$

which implies

$$\frac{t_\epsilon - s_\epsilon}{\epsilon} \leq C_2 - \alpha.$$

Thus we can have

$$-C_2 \leq \frac{t_\epsilon - s_\epsilon}{\epsilon} \leq C_2. \tag{33}$$

Given a unit vector $\mathbf{e} \in \mathbb{R}^d$, let $\eta \in Lip(t_\epsilon, x^\epsilon_{t_\epsilon})$ be given by

$$\eta(t) = \begin{cases} x^\epsilon_{t_\epsilon}(r) & (0 \leq r \leq t_\epsilon), \\ x^\epsilon_{t_\epsilon}(t_\epsilon) + (r - t_\epsilon)\mathbf{e} & (t_\epsilon \leq r \leq T). \end{cases}$$

Since $(t_\epsilon + h, \eta_{t_\epsilon + h}) \in D_k$, we can have from (31) with (33)

$$\begin{aligned} &- C_1 h - C_2 h \\ &\leq \partial_t \phi(t_\epsilon, x^\epsilon(t_\epsilon))h + \langle p_\epsilon, \mathbf{e}\rangle h + h\omega(h; (t_\epsilon, x^\epsilon_{t_\epsilon}), \eta) \\ &\leq C_2 h + \langle p_\epsilon, \mathbf{e}\rangle h + h\omega(h; (t_\epsilon, x^\epsilon_{t_\epsilon}), \eta). \end{aligned}$$

Dividing the above by h and $h \to 0+$, we can obtain

$$-C_1 - 2C_2 \leq \langle p_\epsilon, \mathbf{e}\rangle.$$

Since the unit vector $\mathbf{e}$ is taken arbitrarily, we have

$$|p_\epsilon| \leq C_1 + 2C_2.$$

In a way similar to the above argument, we can have from (32) with (33)

$$|q_\epsilon| \leq C_1 + 2C_2.$$

Since v is a viscosity subsolution, $\phi \in \mathcal{C}^1_u(\mathbf{X}_{0,T})$ and $|\nabla_{x_t} \phi(t_\epsilon, x^\epsilon_{t_\epsilon})| = |p_\epsilon| \leq C_1 + 2C_2$, we have from Definition 5.2

$$-\alpha + \frac{t_\epsilon - s_\epsilon}{\epsilon} + G(p_\epsilon) \geq 0,$$

where $G(p) = \inf_{\beta \in \mathbb{R}^d}\{\langle \beta, p\rangle + L(\beta)\}$. Also, since w is a viscosity supersolution, $\psi \in \mathcal{C}^1_u(\mathbf{X}_{0,T})$ and $|\nabla_{x_t} \psi(s_\epsilon, y^\epsilon_{s_\epsilon})| = |q_\epsilon| \leq C_1 + 2C_2$, we have

$$\alpha + \frac{t_\epsilon - s_\epsilon}{\epsilon} + G(q_\epsilon) \leq 0.$$

Thus we can have

$$2\alpha \leq G(p_\epsilon) - G(q_\epsilon).$$

Take $\beta^*_\epsilon \in \arg\min_{\beta \in \mathbb{R}^d}\{\langle \beta, q_\epsilon\rangle + L(\beta)\}$. Then we can see that

$$\begin{aligned} G(p_\epsilon) - G(q_\epsilon) &\leq \langle \beta^*_\epsilon, p_\epsilon\rangle + L(\beta^*_\epsilon) - \langle \beta^*_\epsilon, q_\epsilon\rangle - L(\beta^*_\epsilon) \\ &\leq |\beta^*_\epsilon||p_\epsilon - q_\epsilon|. \end{aligned}$$

Noting (22) with (30), we can see that there exists a constant $\tilde{M}$ not depending on ϵ such that $|\beta^*_\epsilon| \leq \tilde{M}$. Thus we can have

$$2\alpha \leq \tilde{M}|p_\epsilon - q_\epsilon|.$$

Doing a bit of calculation using the explicit forms of p_ϵ and q_ϵ, we can show that

$$|p_\epsilon - q_\epsilon| \leq \frac{1}{\epsilon}|t_\epsilon - s_\epsilon||x^\epsilon_{t_\epsilon}(t_\epsilon) - y^\epsilon_{s_\epsilon}(s_\epsilon)| + ka(1 + c_k)\frac{|t_\epsilon - s_\epsilon|^2}{\epsilon},$$

where $c_k = \max_{(u, z_u) \in D_k} \|z_u\|_\infty$. Thus we can obtain

$$|p_\epsilon - q_\epsilon| \leq \tilde{c}_k \frac{\nu((t_\epsilon, x^\epsilon_{t_\epsilon}), (s_\epsilon, y^\epsilon_{s_\epsilon}))}{\epsilon},$$

where $\tilde{c}_k = (1/2) + ka(1 + c_k)$. Hence we have

$$2\alpha \leq \tilde{M}\tilde{c}_k \frac{\nu((t_\epsilon, x^\epsilon_{t_\epsilon}), (s_\epsilon, y^\epsilon_{s_\epsilon}))}{\epsilon}.$$

By sending $\epsilon \to 0$, we can have with (28)

$$2\alpha \leq 0.$$

This contradicts the choice of α. Hence $v \leq w$ on $D = \cup_{k=1}^\infty D_k$. ■

Acknowledgement. The authors would like to thank the referees for helpful comments and suggestions.

References

[1] M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity solutions of Hamilton-Jacobi-Bellman Equations*, Birkhäuser, Boston, 1997.

[2] J. Blanchet and H. Lam, State-dependent importance sampling for rare-events simulation: An overview and recent advances, *Surveys in Oper. Res. Manage. Sci.* **17** (2012), 38-59.

[3] G. Buttazzo, M. Giaquinta and S. Hildebrandt, *One-dimensional Variational Problems: An Introduction*, Oxford Univ. Press, New York, 1998.

[4] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Springer, New York, 1998.

[5] B. Dupire, Functional Itô calculus, Bloomberg Portfolio Research Paper, 2009.

[6] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley & Sons, New York, 1997.

[7] P. Dupuis and H. Wang, Importance sampling, large deviations, and differential games, *Stoch. Stoch. Rep.* **76** (2004), 481-508.

[8] P. Dupuis and H. Wang, Subsolutions of an Isaacs equation and efficient schemes for importance sampling, *Math. Oper. Res.* **32** (2007), 723-757.

[9] W.H. Fleming and H.M. Soner, *Controlled Markov Processes and Viscosity Solutions*, 2nd ed., Springer, New York, 2006.

[10] N. Lukoyanov, Hamilton-Jacobi type equation in control problems with hereditary information, *J. App. Math. Mech.* **64** (2000), 243–253

[11] N. Lukoyanov, Functional Hamilton-Jacobi type equations with ci-derivatives in control problems with hereditary information, *Nonlinear Func. Anal. & Appl.* **8** (2003), 535-555.

[12] N. Lukoyanov, On viscosity solution of functional Hamilton-Jacobi type equations for hereditary systems, *Proc. the Steklov Institute of Math.,* Suppl. 2 (2007), 190–200.

[13] H. Kaise, Convergence of discrete-time deterministic games to path-dependent Isaacs partial differential equations with quadratically growing Hamiltonians, *preprint*.

[14] A.V. Kim, *Functional Differential Equations: Applications of i-smooth Calculus*, Kluwer Academic Publishers, Dordrecht, 1999.

[15] Z. Ren, N. Touzi and J. Zhang, An overview of viscosity solutions of path-dependent PDEs, *Stochastic analysis and applications 2014*, eds. D. Crisan, B. Hambly and T. Zariphopoulou (2014), 397–453.

[16] J. Zhang, *Backward Stochastic Differential Equations: From Linear to Fully Nonlinear Theory*, Springer, New York, 2017.