J. Math. Anal. Appl. 497 (2021) 124905

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Singular optimal controls for stochastic recursive systems under convex control constraint

Liangquan Zhang [1]

School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

ARTICLE INFO

Article history:
Received 8 January 2020
Available online 30 December 2020
Submitted by M. Quincampoix

Keywords:
Dynamic programming principle
Forward-backward stochastic differential equations
Malliavin calculus
Maximum principle
Singular optimal controls
Viscosity solution

ABSTRACT

In this paper, we study two kinds of singular optimal controls (SOCs for short) problems where the systems governed by forward-backward stochastic differential equations (FBSDEs for short), in which the control has two components: the regular control, and the singular one. Both drift and diffusion terms may involve the regular control variable. The regular control domain is postulated to be convex. Under certain assumptions, in the framework of the Malliavin calculus, we derive the pointwise second-order necessary conditions for stochastic SOC in the classical sense. This condition is described by two adjoint processes, a maximum condition on the Hamiltonian supported by an illustrative example. A new necessary condition for optimal singular control is obtained as well. Besides, as a by-product, a verification theorem for SOCs is derived via viscosity solutions without involving any derivatives of the value functions. It is worth pointing out that this theorem has wider applicability than the restrictive classical verification theorems. Finally, we focus on the connection between the maximum principle and the dynamic programming principle for such SOCs problem without the assumption that the value function is smooth enough.

1. Introduction

Singular stochastic control problem is a fundamental topic in fields of stochastic control. This problem was first introduced by Bather and Chernoff [7] in 1967 by considering a simplified model for the control of a spaceship. It was then found that there was a connection between the singular control and optimal stopping problem. This link was established through the derivative of the value function of this initial singular control problem and the value function of the corresponding optimal stopping problem. Subsequently, it was considered by Beněs, Shepp, Witzsenhausen (see [9]) and Karatzas and Shreve (see [43–47]).

E-mail address: xiaoquan51011@163.com.
[1] L. Zhang acknowledges the financial support partly by the National Natural Science Foundation of China (Grant No. 11701040, 61871058 & 11871010) and the Fundamental Research Funds for the Central Universities (No. 2019XD-A11).

https://doi.org/10.1016/j.jmaa.2020.124905
0022-247X/

The state process is described by a n-dimensional SDE of the following type:

$$\begin{cases} \mathrm{d}X^{t,x;u,\xi} & = & b\left(s, X^{t,x;u,\xi}(s), u(s)\right)\mathrm{d}s + \sigma\left(s, X^{t,x;u,\xi}(s), u(s)\right)\mathrm{d}W(s) + G(s)\,\mathrm{d}\xi(s), \\ X^{t,x;u,\xi}(t) & = & x, \qquad 0 \leq t \leq s \leq T, \end{cases} \tag{1}$$

on some filtered probability space $(\Omega, \mathcal{F}, P)$, where $b(\cdot,\cdot,\cdot) : [0,T] \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$, $\sigma(\cdot,\cdot,\cdot) : [0,T] \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^{n\times d}$, $G(\cdot) : [0,T] \to \mathbb{R}^{n\times m}$ are given deterministic functions, $(W(s))_{s\geq 0}$ is an d-dimensional Brownian motion, (x,t) are initial time and state, $u(\cdot) : [0,T] \to \mathbb{R}^k$ is a *regular* control process, and $\xi(\cdot) : [0,T] \to \mathbb{R}^m$, with nondecreasing left-continuous with right limits stands for the *singular control*[2] (SC for short). To avoid the risk of confusion, we shall introduce the other definitions of singular control in various senses. Indeed, they are just a coincidence of terminology usage.

The aim is to minimize the cost functional:

$$J(t,x;u,\xi) = \mathbb{E}\left[\int_t^T l\left(s, X^{t,x;u,\xi}(s), u(s)\right)\mathrm{d}s + \int_t^T K(s)\,\mathrm{d}\xi(s)\right], \tag{2}$$

where

$$l(\cdot,\cdot,\cdot) : [0,T] \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R},$$
$$K(\cdot) : [0,T] \to \mathbb{R}^m_+ \triangleq \{x \in \mathbb{R}^m : x_i \geq 0,\ i = 1, \ldots m\}$$

are given deterministic functions, where $l(\cdot)$ represents the running cost tare of the problem and K the cost rate of applying the singular control.

We mention that there are four approaches to deal with singular control: The first, partial differential equations (PDE for short) and on variational arguments, can be found in the works of Alvarez [1,2], Chow, Menaldi, and Robin [23], Karatzas [44], Karatzas and Shreve [47], and Menaldi and Taksar [57]. The second one is related to probabilistic methods; see Baldursson [5], Boetius [12,13], Boetius and Kohlmann [14], El Karoui and Karatzas [28,29], Karatzas [43], and Karatzas and Shreve [45,46]. Third, the DPP, has been studied in a general context, for example, by Boetius [13], Haussmann and Suo [39], Fleming and Soner [32] and Zhang [79]. At last the maximum principle for optimal singular controls (see, for example, Cadenillas and Haussmann [20], Dufour and Miller [27], Dahl and Øksendal [25] see references therein).

Singular controls are used in diverse fields such as mathematical finance (see Baldursson and Karatzas [6], Chiarolla, Haussmann [22], Kobila [50], Karatzas, Wang [48], Davis, Norman [26] and Pagès and Possamaï [62]), manufacturing systems (see Shreve, Lehoczky, and Gaver [70]), and queuing systems (see Martins and Kushner [56]).

Completely different from the singular control introduced above, to the best of our knowledge, there are two other types of singular optimal controls, in which the first-order necessary conditions turn out to be trivial. We list briefly as follows:

- *Singular optimal control in the classical sense (SOCCS for short)*, is the optimal control for which the gradient and the Hessian of the corresponding Hamiltonian with respect to the control variable vanish/degenerate.
- *Singular optimal control in the sense of Pontryagin-type maximum principle (SOCSPMP for short)*, is the optimal control for which the corresponding Hamiltonian is equal to a constant in the control region.

[2] Because the measure $\mathrm{d}\xi_s$ may be singular with respect to the Lebesgue measure $\mathrm{d}s$.

When an optimal control is singular in certain senses above (SOCCS and SOCSPMP), usually the first-order necessary condition could not carry sufficient information for the further theoretical analysis and numerical computation, and consequently it is necessary to investigate the second order necessary conditions. In the deterministic setting, reader can refer many articles in this direction (see [8,33,35,36,49,52,54] and references therein).

As for the second-order necessary conditions for stochastic singular optimal controls (SOCCS and SOCSPMP), there are some work should be mentioned, for instance [80,81] (note that singular control $\xi(\cdot)$ in these articles does not appear in systems). Tang [71] obtained a pointwise second order maximum principle for stochastic singular optimal controls in the sense of the Pontryagin-type maximum principle whenever the control variable u does not enter into the diffusion term. Meanwhile, Tang addressed an integral-type second-order necessary condition for stochastic optimal controls with convex control constraints. Zhang and Zhang [80] also establish certain pointwise second-order necessary conditions for stochastic singular (SOCCS) optimal controls, in which both drift and diffusion terms in may depend on the control variable u with convex control region U by making use of Malliavin calculus technique. Later, adopting the same idea but with large complicated analysis, Zhang et al. [81] deepen this research for the general case when the control region is nonconvex.

The theory of *backward stochastic differential equation* (BSDE for short) can be traced back to Bismut [10,11] who studied linear BSDE motivated by stochastic control problems. Pardoux and Peng 1990 [63] proved the well-posedness for nonlinear BSDE. Duffie and Epstein (1992) introduced the notion of recursive utilities in continuous time, which is actually a type of BSDE where the generator f is independent of z. El Karoui et al. (1997, 2001) extended the recursive utility to the case where f contains z. The term z can be interpreted as an ambiguity aversion term in the market (see Chen and Epstein 2002 [21]). Particularly, the celebrated Black-Scholes formula indeed provided an effective way of representing the option price (which is the solution to a kind of linear BSDE) through the solution to the Black-Scholes equation (parabolic partial differential equation actually). Since then, BSDE has been extensively studied and used in the areas of applied probability and optimal stochastic controls, particularly in financial engineering (cf. for instance [30]).

By means of BSDE, Peng (1990) [64] considered the following type of stochastic optimal control problem: minimize a cost function

$$J(u(\cdot)) = \mathbb{E}\left[\int_0^T l(x(t), u(t))\,\mathrm{d}t + h(x(T))\right],$$

subject to

$$\left\{\begin{array}{rcl} \mathrm{d}x(t) &=& g(t, x(t), u(t))\,\mathrm{d}t + \sigma(t, x(t), u(t))\,\mathrm{d}W(t), \\ x(0) &=& x_0, \end{array}\right. \tag{3}$$

over an admissible control domain which need not be convex, and the diffusion coefficients depend on the control variable. In his paper, by spike variational method and the second order adjoint equations, Peng [64] obtained a general stochastic maximum principle for the above optimal control problem. It was just the adjoint equations in stochastic optimal control problems that motivated the famous theory of BSDE (cf. [63]).

Later, Peng first [66] studied a stochastic optimal control problem where state variables are described by the system of FBSDEs:

$$\left\{\begin{array}{rcl} \mathrm{d}x(t) &=& f(t, x(t), u(t))\,\mathrm{d}t + \sigma(t, x(t), u(t))\,\mathrm{d}W(t), \\ \mathrm{d}y(t) &=& g(t, x(t), u(t))\,\mathrm{d}t + z(t)\,\mathrm{d}W(t), \\ x(0) &=& x_0, y(T) = y, \end{array}\right. \tag{4}$$

where x and y are given deterministic constants. The optimal control problem is to minimize the cost function:

$$J(u(\cdot)) = \mathbb{E}\left[\int_0^T l(t, x(t), y(t), u(t))\,\mathrm{d}t + h(x(T)) + \gamma(y(0))\right],$$

over an admissible control domain which is convex. Later, Xu [75] studied the following non-fully coupled forward-backward stochastic control system:

$$\begin{cases} \mathrm{d}x(t) &= f(t, x(t), u(t))\,\mathrm{d}t + \sigma(t, x(t))\,\mathrm{d}W(t), \\ \mathrm{d}y(t) &= g(t, x(t), y(t), z(t), u(t))\,\mathrm{d}t + z(t)\,\mathrm{d}W(t), \\ x(0) &= x_0, y(T) = h(x(T)). \end{cases} \tag{5}$$

The optimal control problem is to minimize the cost function $J\left(u_{(\cdot)}\right) = \mathbb{E}\gamma(y_0)$, over $\mathcal{U}_{ad}$, but the control domain is non-convex. Wu [73] firstly gave the maximum principle for optimal control problem of fully coupled forward-backward stochastic system:

$$\begin{cases} \mathrm{d}x(t) &= f(t, x(t), y(t), z(t), u(t))\,\mathrm{d}t + \sigma(t, x(t), y(t), z(t), u(t))\,\mathrm{d}W(t), \\ \mathrm{d}y(t) &= -g(t, x(t), y(t), z(t), u(t))\,\mathrm{d}t + z(t)\,\mathrm{d}W(t), \\ x(0) &= x_0, \qquad y(T) = \xi, \end{cases} \tag{6}$$

where ξ is a random variable and the cost function:

$$J(u(\cdot)) = \mathbb{E}\left[\int_0^T L(t, x(t), y(t), z(t), u(t))\,\mathrm{d}t + \Phi(x(T)) + h(y(0))\right].$$

The optimal control problem is to minimize the cost function $J\left(v_{(\cdot)}\right)$ over an admissible control domain which is convex. Ji and Zhou [42] obtained a maximum principle for stochastic optimal control of non-fully coupled forward-backward stochastic system with terminal state constraints. Shi and Wu [69] studied the maximum principle for fully coupled forward-backward stochastic system:

$$\begin{cases} \mathrm{d}x(t) &= b(t, x(t), y(t), z(t), u(t))\,\mathrm{d}t + \sigma(t, x(t), y(t), z(t))\,\mathrm{d}B_t, \\ \mathrm{d}y(t) &= -f(t, x(t), y(t), z(t), u(t))\,\mathrm{d}t + z(t)\,\mathrm{d}B_t, \\ x(0) &= x_0, \quad y(T) = h(x(T)), \end{cases} \tag{7}$$

and the cost function is

$$J(u(\cdot)) = \mathbb{E}\left[\int_0^T l(t, x(t), y(t), z(t), u(t))\,\mathrm{d}t + \Phi(x(T)) + \gamma(y(0))\right].$$

The control domain is non-convex but the forward diffusion does not contain the control variable.

Subsequently, in order to study the backward linear-quadratic optimal control problem, Kohlmann and Zhou [51], Lim and Zhou [55] developed a new method for handling this problem. The term z is regarded as a control process and the terminal condition $y(T) = h(x(T))$ as a constraint, and then it is possible to use the Ekeland variational principle to obtain the maximum principle. Adopting this idea, Yong [76] and Wu [74] independently established the maximum principle for the recursive stochastic optimal control problem (noting the diffusion term containing control variable with non-convex control region). Nonetheless, the maximum principle derived by these methods involves two unknown parameters. Therefore, the hard questions raise as follows: What is the second-order variational equation for the BSDE? How to obtain the second-order adjoint equation since the quadratic form with respect to the variation of z. All of which seem to be extremely complicated.

Hu [40] overcomes the above difficulties by introducing two new adjoint equations. Then, the second-order variational equation for the BSDE and the maximum principle are obtained. The main difference of his variational equations with those in Peng [64] consists in the term $\langle p(t), \delta\sigma(t)\rangle I_{E_\varepsilon}(t)$ in the variation of z. Due to the term $\langle p(t), \delta\sigma(t)\rangle I_{E_\varepsilon}(t)$ in the variation of z, Hu obtained a global maximum principle which is novel and different from that in Wu [74], Yong [76] and previous work, which solves completely Peng's open problem. Furthermore, Hu's maximum principle is stronger than the one in Wu [74], Yong [76]. For a general case, reader can refer [41].

Motivated by above work, in this paper, we consider singular controls problem of the following type:

$$\begin{cases} \mathrm{d}X^{t,x;u,\xi}(s) &= b\left(s, X^{t,x;u,\xi}(s), u(s)\right)\mathrm{d}s + \sigma\left(s, X^{t,x;u,\xi}(s), u(s)\right)\mathrm{d}W(s) + G(s)\,\mathrm{d}\xi(s), \\ \mathrm{d}Y^{t,x;u,\xi}(s) &= -f\left(t, X^{t,x;u,\xi}(s), Y^{t,x;u,\xi}(s), Z^{t,x;u,\xi}(s), u(s)\right)\mathrm{d}s \\ & \quad + Z^{t,x;u,\xi}(s)\,\mathrm{d}W(s) - K\mathrm{d}\xi(s), \\ X^{t,x;u,\xi}(t) &= x,\ Y^{t,x;u,\xi}(T) = \Phi\left(X^{t,x;u,\xi}(T)\right), \qquad 0 \le t \le s \le T, \end{cases} \tag{8}$$

with the similar cost functional

$$J(t,x;u,\xi) = Y^{t,x;u,\xi}(s)\big|_{s=t}. \tag{9}$$

Wang [72] firstly introduced and studied a class of singular control problems with recursive utility, where the cost function is determined by BSDE. Under certain assumptions, the author proved that the value function is a nonnegative, convex solution of the H-J-B equation. However, FBSDEs in Wang [72] do not contain the regular control and the generator is not general case. In our work, using some properties of the BSDE and analysis technique, we expand the extension of the MP for SOC to the recursive control problem in Zhang and Zhang [80]. To the best of our knowledge, such singular optimal controls problems of FBSDEs (8) via two kinds of singular controls have not been explored before. We shall establish some pointwise second-order necessary conditions for stochastic optimal controls of FBSDEs. Both drift and diffusion terms may contain the control variable u, and we assume that the control region U is convex. We also consider the pointwise second-order necessary condition, which is easier to verify in practical applications.

As claimed in [80], quite different from the deterministic setting, there exist some essential difficulties in deriving the pointwise second-order necessary condition from an integral-type one whenever the diffusion term depends on the control variable, even for the case of convex control domain. We overcome these difficulties by means of some technique from the Malliavin calculus. For general case, namely, the control region is non-convex can be found in [81].

In this paper, we are interested in studying singular optimal controls for FBSDEs (8). Compared with above literature, our paper has several new features. The novelty of the formulation and the contribution in this paper may be stated as follows:

- Our control systems in this paper are governed by FBSDEs which exactly extends the work of Zhang and Zhang [80] to utilities. Our work is the first time to establish the pointwise second order necessary condition for stochastic singular optimal control in the classical sense for FBSDEs, a new necessary condition for singular control is involved as well. In this sense, our paper actually considers two kinds of singular controls problems simultaneously, which is interesting to deepen this research.
- We derive a new verification theorem for optimal singular controls via viscosity solution, which responses to the question raised in Zhang [79]; Meanwhile, we study the relationship between the adjoint equations

derived and value function, which extends the smooth case considered by Cadenillas and Haussmann [20] to the framework of viscosity solution for stochastic recursive systems.

The rest of this paper is organized as follows: after some preliminaries in the second section, we are devoted the third section to the MP for two kinds of singular optimal controls. A concrete example is concluded with as well. Then, in Section 4, we study the verification theorem for singular optimal controls via viscosity solutions. Finally, we establish the relationship between the DPP and MP for viscosity solution. Some proofs of lemmas are displayed in Appendix A.

2. Preliminaries and notations

Throughout this paper, we denote by $\mathbb{R}^n$ the space of n-dimensional Euclidean space, by $\mathbb{R}^{n\times d}$ the space the matrices with order $n \times d$. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ be a complete filtered probability space on which a *one*-dimensional standard Brownian motion $W(\cdot)$ is defined, with $\{\mathcal{F}_t\}_{t\geq 0}$ being its natural filtration, augmented by all the P-null sets. Given a subset U (nonempty, bounded, and convex) of $\mathbb{R}^k$, we will denote $\mathcal{U}[0,T] = \mathcal{U}_1 \times \mathcal{U}_2$, separately, the class of measurable, adapted processes $(u, \xi) : [0,T] \times \Omega \to U \times [0,\infty)^m$, with ξ nondecreasing left-continuous with right limits and $\xi_0 = 0$, moreover, $\mathbb{E}\left[\sup\limits_{0\leq t\leq T} |u(t)|^2 + |\xi(T)|^2\right] < \infty$. ξ is called *singular control*. For each $t > 0$, we denote by $\{\mathcal{F}_s^t, t \leq s \leq T\}$ the natural filtration of the Brownian motion $\{W(s) - W(t)\}_{t\leq s\leq T}$, augmented by the P-null sets of $\mathcal{F}$. $\top$ appearing as superscript denotes the transpose of a matrix. In what follows, C represents a generic constant, which can be different from line to line.

We now introduce the following spaces of processes:

$$\mathcal{S}^2(0,T;\mathbb{R}) \triangleq \left\{\mathbb{R}^n\text{-valued } \mathcal{F}_t\text{-adapted process } \phi(t);\ \mathbb{E}\left[\sup_{0\leq t\leq T} |\phi(t)|^2\right] < \infty\right\},$$

$$\mathcal{M}^2(0,T;\mathbb{R}) \triangleq \left\{\mathbb{R}^n\text{-valued } \mathcal{F}_t\text{-adapted process } \varphi(t);\ \mathbb{E}\left[\int_0^T |\varphi(t)|^2 \, \mathrm{d}t\right] < \infty\right\},$$

and denote $\mathcal{N}^2[0,T] = \mathcal{S}^2(0,T;\mathbb{R}^n) \times \mathcal{S}^2(0,T;\mathbb{R}) \times \mathcal{M}^2(0,T;\mathbb{R}^n)$. Clearly, $\mathcal{N}^2[0,T]$ forms a Banach space.

For any $u(\cdot) \times \xi(\cdot) \in \mathcal{U}_1 \times \mathcal{U}_2$, we study the stochastic control systems governed by FBSDEs (8).

We assume that the following conditions hold:

(A1) The coefficients $b : [0,T] \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$, $\sigma : [0,T] \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$, are twice continuously differentiable with respect to x; b, b_x, b_{xx}, σ, σ_x, σ_{xx} are continuous in (x,u); b_x, b_{xx}, σ_x, σ_{xx} are bounded b, σ are bounded by $C(1 + |x| + |u|)$ for some positive constant C. Moreover, for any (t, x_1, u_1), $(t, x_2, u_2) \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^k$,

$$\begin{aligned} |b(t,0,x)| + |\sigma(t,0,u)| &\leq C, \\ \left|b_{(x,u)^2}(t,x_1,u_1) - b_{(x,u)^2}(t,x_2,u_2)\right| &\leq C(|x_1 - x_2| + |u_1 - u_2|), \\ \left|\sigma_{(x,u)^2}(t,x_1,u_1) - \sigma_{(x,u)^2}(t,x_2,u_2)\right| &\leq C(|x_1 - x_2| + |u_1 - u_2|). \end{aligned}$$

(A2) The coefficients $f : [0,T] \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}$, $\Phi : \mathbb{R}^n \to \mathbb{R}$, are twice continuously differentiable with respect to (x,y,z). K is a given deterministic matrix. f, Df, $D^2 f$ are continuous in (x,y,z,u).

There exists constant $C>0$ such that for any (t,x_1,y_1,z_1,u_1), $(t,x_2,y_2,z_2,u_2)\in[0,T]\times\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}^k$,

$$|f(t,x,y,z,u)|\leq C(1+|x|+|y|+|z|),$$
$$|f_x(t,x,y,z,u)|+|f_y(t,x,y,z,u)|+|f_z(t,x,y,z,u)|+|f_u(t,x,y,z,u)|\leq C,$$
$$\begin{aligned}&|f_{xx}(t,x,y,z,u)|+|f_{xu}(t,x,y,z,u)|+|f_{yu}(t,x,y,z,u)|\\&\quad+|f_{yy}(t,x,y,z,u)|+|f_{zz}(t,x,y,z,u)|+|f_{zu}(t,x,y,z,u)|\\&\quad+|f_{uu}(t,x,y,z,u)|\leq C,\end{aligned}$$
$$\begin{aligned}&\left|f_{(x,y,z,u)^2}(t,x_1,y_1,z_1,u_1)-f_{(x,y,z,u)^2}(t,x_2,y_2,z_2,u_2)\right|\\&\qquad\leq C(|x_1-x_2|+|y_1-y_2|+|z_1-z_2|+|u_1-u_2|),\end{aligned}$$

and

$$\begin{aligned}&\Phi(x)\leq C\left(1+|x|^2\right),\ \Phi_x(x)\leq C(1+|x|),\\&\Phi_{xx}(x)\leq C,\ |\Phi_{xx}(x_1)-\Phi_{xx}(x_2)|\leq C|x_1-x_2|.\end{aligned}$$

Under above assumptions (A1)-(A2), for any $u(\cdot)\times\xi(\cdot)\in\mathcal{U}_1\times\mathcal{U}_2$, it is easy to check that FBSDEs (8) admit a unique $\mathcal{F}_t$-adapted solution denoted by the triple $(X_\cdot^{t,x;u,\xi},Y_\cdot^{t,x;u,\xi},Z_\cdot^{t,x;u,\xi})\in\mathcal{N}^2[0,T]$ (see Pardoux and Peng [63]).

Like Peng [67], given any control processes $u(\cdot)\times\xi(\cdot)\in\mathcal{U}_1\times\mathcal{U}_2$, we introduce the following cost functional:

$$J(t,x;u(\cdot),\xi(\cdot))=Y^{t,x;u,\xi}(s)\big|_{s=t},\qquad(t,x)\in[0,T]\times\mathbb{R}^n.\tag{10}$$

We are interested in the *value function* of the stochastic optimal control problem:

$$\begin{aligned}V(t,x)&=J(t,x;\hat{u}(\cdot),\hat{\xi}(\cdot))\\&=ess\inf_{u(\cdot)\times\xi(\cdot)\in\mathcal{U}_1\times\mathcal{U}_2}J(t,x;u(\cdot),\xi(\cdot)),\ (t,x)\in[0,T]\times\mathbb{R}^n.\end{aligned}\tag{11}$$

Since the value function (11) is defined by the solution of controlled FBSDEs (8), so from the existence and uniqueness, V is *well-defined.*

The following estimate is very useful whose proof can be found in Briand et al. 2003 [16].

Lemma 1. *Let $\left(y^i,z^i\right)$, $i=1,2$, be the solution to the following*

$$y^i(t)=\xi^i+\int_t^T f^i\left(s,y^i(s),z^i(s)\right)\mathrm{d}s-\int_t^T z^i(s)\,\mathrm{d}W(s),\tag{12}$$

where $\xi^i\in\mathcal{F}_T$ and $\mathbb{E}\left[\left|\xi^i\right|^\beta\right]<\infty$, whilst $f^i\left(s,y^i,z^i\right)$ satisfies the conditions (A2), *and*

$$\mathbb{E}\left[\left(\int_t^T\left|f^i\left(s,y^i(s),z^i(s)\right)\right|\mathrm{d}s\right)^\beta\right]<\infty;$$

Then, for some $\beta\geq2$, there exists a positive constant C_β such that

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|y^1(t)-y^2(t)\right|^\beta+\left(\int_0^T\left|z^1(s)-z^2(s)\right|^2\mathrm{d}s\right)^{\frac{\beta}{2}}\right]$$

$$\le C_\beta\mathbb{E}\left[\left|\xi^1-\xi^2\right|^\beta+\left(\int_t^T\left|f^1\left(s,y^1(s),z^1(s)\right)-f^2\left(s,y^2(s),z^2(s)\right)\right|\mathrm{d}s\right)^\beta\right].$$

Particularly, whenever putting $\xi^2=0$*, and* $f^2=0$*, one has*

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|y^1(t)\right|^\beta+\left(\int_0^T\left|z^1(s)\right|^2\mathrm{d}s\right)^{\frac{\beta}{2}}\right]\le C_\beta\mathbb{E}\left[\left|\xi^1\right|^\beta+\left(\int_t^T\left|f^1(s,0,0)\right|\mathrm{d}s\right)^\beta\right].$$

Now let us recall briefly the notion of differentiation on Wiener space (see the expository papers by Nualart 1995 [59], Nualart and Pardoux [60] and Ocone 1988 [61]).

- $C_b^k\left(\mathbb{R}^k,\mathbb{R}^q\right)$ will denote the set of functions of class C^k from $\mathbb{R}^k$ into $\mathbb{R}^q$ whose partial derivatives of order less than or equal to k are bounded.
- Let $\mathcal{S}$ denote the set of random variables ξ of the form $\xi=\varphi(W\left(h^1\right),W\left(h^2\right),\cdots,W\left(h^k\right))$, where $\varphi\in C_b^\infty\left(\mathbb{R}^k,\mathbb{R}\right)$, $h^1,h^2,\cdots h^k\in L^2\left([0,T];\mathbb{R}^n\right)$, and $W\left(h^i\right)=\int_0^T\left\langle h_s^i,\mathrm{d}W(s)\right\rangle$.
- If $\xi\in\mathcal{S}$ is of the above form, we define its derivative as being the n-dimensional process

$$\mathcal{D}_\theta\xi=\sum_{j=1}^k\frac{\partial}{\partial x_j}\varphi\left(W\left(h^1\right),W\left(h^2\right),\cdots,W\left(h^k\right)\right)h_\theta^j,\ 0\le\theta\le T.$$

For $\xi\in\mathcal{S}$, $p>1$, we define the norm

$$\|\xi\|_{1,p}=\left[E\left\{|\xi|^p+\left(\int_0^T\left|\mathcal{D}_\theta\xi\right|^2\mathrm{d}\theta\right)^{\frac{p}{2}}\right\}\right]^{\frac{1}{p}}.$$

It can be shown (Nualart 1995) that the operator $\mathcal{D}$ has a closed extension to the space $\mathbb{D}^{1,p}$, the closure of $\mathcal{S}$ with respect to the norm $\|\cdot\|_{1,p}$. Observe that if ξ is $\mathcal{F}_t$-measurable, then $\mathcal{D}_\theta\xi=0$ for $\theta\in(t,T]$. We denote by $\mathcal{D}_\theta^i\xi$, the ith component of $\mathcal{D}_\theta\xi$.

Let $\mathbb{L}^{1,p}\left(\mathbb{R}^d\right)$ denote the set of $\mathbb{R}^d$-valued progressively measurable processes $\{u(t,\omega),0\le t\le T;\omega\in\Omega\}$ such that

- For a.e. $t\in[0,T]$, $u(t,\cdot)\in\mathbb{D}^{1,p}\left(\mathbb{R}^n\right)$;
- $(t,\omega)\to\mathcal{D}u(t,\omega)\in\left(L^2([0,T])\right)^{n\times d}$ admits a progressively measurable version;
- We have

$$\|u\|_{1,p}=\mathbb{E}\left[\left(\int_0^T|u(t)|^2\mathrm{d}t\right)^{\frac{p}{2}}+\left(\int_0^T\int_0^T\left|\mathcal{D}_\theta u(t)\right|^2\mathrm{d}\theta\mathrm{d}t\right)^{\frac{p}{2}}\right]<+\infty.$$

Note that for each (θ,t,ω), $\mathcal{D}_\theta u(t,\omega)$ is an $n\times d$ matrix. Hence, $\left|\mathcal{D}_\theta u(t)\right|^2=\sum_{i,j}\left|\mathcal{D}_\theta^i u_j(t)\right|^2$. Obviously, $\mathcal{D}_\theta u(t,\omega)$ is defined uniquely up to sets of $\mathrm{d}\theta\otimes\mathrm{d}t\otimes\mathrm{d}P$ measure zero. Moreover, denote by $\mathbb{L}_{\mathbb{F}}^{1,p}\left(\mathbb{R}^d\right)$ the set of all adapted processes in $\mathbb{L}^{1,p}\left(\mathbb{R}^d\right)$.

We define the following notations from Zhang and Zhang [80]:

$$\begin{aligned}\mathbb{L}_{2+}^{1,p}\left(\mathbb{R}^d\right) \quad := \quad & \Big\{ \varphi(\cdot) \in \mathbb{L}^{1,p}\left(\mathbb{R}^d\right) \Big| \, \exists \mathcal{D}^{+}\varphi(\cdot) \in L^2\left([0,T]\times\Omega;\mathbb{R}^n\right) \text{ such that} \\ & f_\varepsilon(s) := \sup_{s<t<(s+t)\wedge T} \mathbb{E}\left|\mathcal{D}_s\varphi(t) - \mathcal{D}^{+}\varphi(s)\right|^2 < \infty, \text{ a.e. } s\in[0,T], \\ & f_\varepsilon(s) \text{ is measurable on } [0,T] \text{ for any } \varepsilon>0, \text{ and } \lim_{\varepsilon\to 0^+}\int_0^T f_\varepsilon(s)\,\mathrm{d}s = 0\Big\};\end{aligned}$$

and

$$\begin{aligned}\mathbb{L}_{2-}^{1,p}\left(\mathbb{R}^d\right) \quad := \quad & \Big\{ \varphi(\cdot) \in \mathbb{L}^{1,p}\left(\mathbb{R}^d\right) \Big| \, \exists \mathcal{D}^{-}\varphi(\cdot) \in L^2\left([0,T]\times\Omega;\mathbb{R}^n\right) \text{ such that} \\ & g_\varepsilon(s) := \sup_{(s-\varepsilon)\vee 0<t<s} \mathbb{E}\left|\mathcal{D}_s\varphi(t) - \mathcal{D}^{-}\varphi(s)\right|^2 < \infty, \text{ a.e. } s\in[0,T], \\ & g_\varepsilon(s) \text{ is measurable on } [0,T] \text{ for any } \varepsilon>0, \text{ and } \lim_{\varepsilon\to 0^+}\int_0^T g_\varepsilon(s)\,\mathrm{d}s = 0\Big\}.\end{aligned}$$

Denote $\mathbb{L}_2^{1,p}\left(\mathbb{R}^d\right) = \mathbb{L}_{2+}^{1,p}\left(\mathbb{R}^d\right) \cap \mathbb{L}_{2-}^{1,p}\left(\mathbb{R}^d\right)$. For any $\varphi(\cdot) \in \mathbb{L}_2^{1,p}\left(\mathbb{R}^d\right)$, denote $\nabla\varphi(\cdot) = \mathcal{D}^{+}\varphi(\cdot) + \mathcal{D}^{-}\varphi(\cdot)$. Whenever φ is adapted, it follows that $\mathcal{D}_s\varphi(t) = 0$ for $t<s$. Furthermore, $\nabla\varphi(\cdot) = \mathcal{D}^{+}\varphi(\cdot)$ since $\mathcal{D}^{-}\varphi(\cdot) = 0$. Put $\mathbb{L}_{2,\mathbb{F}}^{1,p}\left(\mathbb{R}^d\right)$ as the set of all adapted processes in $\mathbb{L}_2^{1,p}\left(\mathbb{R}^d\right)$.

3. Maximum principle of singular optimal controls

This section will study the optimal controls separately. Due to the special structure of control systems, we shall first consider the singular control part, deriving the necessary condition, subsequently, regular part. The initial condition will be fixed to be $(0,x)$, $x\in\mathbb{R}^n$. At the beginning let us suppose that $\left(\bar{u}(\cdot), \bar{\xi}(\cdot)\right) \in \mathcal{U}_1 \times \mathcal{U}_2$ is an optimal control and denote by $\left(X^{0,x;\bar{u},\bar{\xi}}(\cdot), Y^{0,x;\bar{u},\bar{\xi}}(\cdot), Z^{0,x;\bar{u},\bar{\xi}}(\cdot)\right)$ the optimal solution of (8). Our maximum principle will be proved in two steps. The first variational inequality is derived from the fact

$$J\left(0,x,u^{\varepsilon}(\cdot),\bar{\xi}(\cdot)\right) - J\left(0,x,\bar{u}(\cdot),\bar{\xi}(\cdot)\right) \geq 0, \tag{13}$$

where $u^{\varepsilon}(\cdot)$ is a convex perturbation of optimal control. The second variational inequity is attained from the inequity

$$J\left(0,x,\bar{u}(\cdot),\xi^{\varepsilon}(\cdot)\right) - J\left(0,x,\bar{u}(\cdot),\bar{\xi}(\cdot)\right) \geq 0, \tag{14}$$

where $\xi^{\varepsilon}(\cdot)$ is a convex perturbation of ξ.

3.1. Optimal singular control

For $l = b(\cdot)$, $\sigma(\cdot)$, $f(\cdot)$, we denote

$$\begin{aligned}\bar{l}_x(r,\cdot) &= l_x\left(r, X^{0,x;\bar{u},\bar{\eta}}(r), Y^{0,x;\bar{u},\bar{\eta}}(r), Z^{0,x;\bar{u},\bar{\eta}}(r), \bar{u}(r)\right), \\ \bar{l}_y(r,\cdot) &= l_y\left(r, X^{0,x;\bar{u},\bar{\eta}}(r), Y^{0,x;\bar{u},\bar{\eta}}(r), Z^{0,x;\bar{u},\bar{\eta}}(r), \bar{u}(r)\right).\end{aligned}$$

Let us introduce the following

Proposition 2. *Let* (A1)-(A2) *hold, and let* $\left(X^{0,x;\bar{u},\bar{\eta}}(\cdot), Y^{0,x;\bar{u},\bar{\eta}}(\cdot), Z^{0,x;\bar{u},\bar{\eta}}(\cdot)\right) \in \mathcal{N}^2(0,T;\mathbb{R}^n)$ *be an optimal solution. Then, the following FBSDEs:*

$$\begin{cases} \mathrm{d}\mathfrak{p}(r) &= -\left[\bar{b}_x(r,\cdot)^{\top}\mathfrak{p}(r) + \bar{\sigma}_x(r,\cdot)^{\top}\mathfrak{k}(r) - \bar{f}_x(r,\cdot)^{\top}\mathfrak{q}(r)\right]\mathrm{d}r + \mathfrak{k}(r)\,\mathrm{d}W(r), \\ \mathrm{d}\mathfrak{q}(r) &= \bar{f}_y(r,\cdot)^{\top}\mathfrak{q}(r)\,\mathrm{d}r + \bar{f}_z(r,\cdot)^{\top}\mathfrak{q}(r)\,\mathrm{d}W(r), \\ \mathfrak{p}(T) &= -\Phi_x\left(X^{t,x;u^{\varepsilon}}(T)\right)\mathfrak{q}(T), \qquad \mathfrak{q}(0) = 1, \end{cases} \tag{15}$$

admit an adapted solution $(\mathfrak{p}(\cdot), \mathfrak{q}(\cdot), \mathfrak{k}(\cdot)) \in \mathcal{N}^2(0,T;\mathbb{R}^n)$.

Theorem 3. *Let* (A1)-(A2) *hold. If* $\left(X^{\bar{u},\bar{\eta}}(\cdot), Y^{\bar{u},\bar{\eta}}(\cdot), Z^{\bar{u},\bar{\eta}}(\cdot), \bar{u}(\cdot), \bar{\xi}(\cdot)\right)$ *is an optimal solution of (8), then there exists a unique pair of adapted processes* $(\mathfrak{p}(\cdot), \mathfrak{q}(\cdot))$ *satisfying (15) such that*

$$P\left\{\mathfrak{q}(t) K_{(i)} - \mathfrak{p}^\top(t) G_{(i)}(t) \geq 0,\ t \in [0,T],\ \forall i\right\} = 1, \tag{16}$$

and

$$P\left\{\sum_{i=1}^{m} \int_0^T I_{\{\mathfrak{q}(r)K_{(i)} - \mathfrak{p}^T(r)G_{(i)}(t) > 0\}} \mathrm{d}\bar{\xi}^{(i)}(r) = 0\right\} = 1. \tag{17}$$

Before the proof, we need some lemmas. At the beginning, we introduce the convex perturbation

$$(\bar{u}(t), \xi^\alpha(t)) = \left(\bar{u}(t), \bar{\xi}(t) + \alpha\left(\xi(t) - \bar{\xi}(t)\right)\right),$$

where $\alpha \in [0,1]$ and $\xi(\cdot)$ is an arbitrary element of $\mathcal{U}_2$. We now introduce the following variational equations of (8):

$$\begin{cases} \mathrm{d}x^1(t) &=& \bar{b}_x(t) x^1(t)\,\mathrm{d}t + \bar{\sigma}_x(t) x^1(t)\,\mathrm{d}W(t) + G(t)\,\mathrm{d}\left(\xi(t) - \bar{\xi}(t)\right), \\ \mathrm{d}y^1(t) &=& -\bar{f}_x(t) x^1(t) - \bar{f}_y(t) y^1(t) - \bar{f}_z(t) z^1(t)\,\mathrm{d}t + z^1(t)\,\mathrm{d}W(t) \\ && -K\mathrm{d}\left(\xi(t) - \bar{\xi}(t)\right), \\ y^1(T) &=& \Phi_x\left(X^{0,x;\bar{u},\bar{\xi}}(T)\right) x^1(T),\ x^1(0) = 0. \end{cases} \tag{18}$$

From (A1)-(A2) it is easy to check that (18) has a unique strong solution. Moreover, we have

Lemma 4. *Under the Assumptions* (A1)-(A2)*, we have*

$$\lim_{\alpha\to 0} \mathbb{E}\left[\left|\frac{X^{0,x;\bar{u},\xi^\alpha}(t) - X^{0,x;\bar{u},\bar{\xi}}(t)}{\alpha} - x^1(t)\right|^2\right] = 0,\ t \in [0,T], \tag{19}$$

$$\lim_{\alpha\to 0} \mathbb{E}\left[\left|\frac{Y^{0,x;\bar{u},\xi^\alpha}(t) - Y^{0,x;\bar{u},\bar{\xi}}(t)}{\alpha} - y^1(t)\right|^2\right] = 0,\ t \in [0,T], \tag{20}$$

$$\lim_{\alpha\to 0} \mathbb{E}\left[\int_0^T \left|\frac{Z^{0,x;\bar{u},\xi^\alpha}(t) - Z^{0,x;\bar{u},\bar{\xi}}(t)}{\alpha} - z^1(t)\right|^2 \mathrm{d}t\right] = 0. \tag{21}$$

The proof can be seen in the Appendix.

Proof of Theorem 3. Applying Itô's formula to $\left\langle \mathfrak{p}(\cdot), x^1(\cdot)\right\rangle + \mathfrak{q}(\cdot) y^1(\cdot)$ on $[0,T]$ yields

$$0 \leq y^1(0) = \mathbb{E}\left[\int_0^T \left(\mathfrak{q}(r) K - \mathfrak{p}^\top(r) G(t)\right) \mathrm{d}\left(\xi(t) - \bar{\xi}(t)\right) \mathrm{d}t\right]. \tag{22}$$

In particular, let $\xi \in \mathcal{U}_2$ be a process satisfying $P\left\{\sum_i \int_0^T G(s)\,\mathrm{d}\xi^{(i)}(s) < \infty\right\}$ and such that (22) and

$$\mathrm{d}\xi^{(i)}(s) = \begin{cases} 0 & \text{if } \mathfrak{q}(r) K_{(i)} - \mathfrak{p}^{\top}(r) G_{(i)}(t) > 0, \\ \hat{\xi}^{(i)}(s) & \text{otherwise,} \end{cases}$$

holds where $\xi^{(i)}(s)$ denotes the ith component. Then,

$$\begin{aligned} &\mathbb{E}\left[\sum_{i=1}^{m}\int_{0}^{T}\left(\mathfrak{q}(r) K_{(i)} - \mathfrak{p}^{\top}(r) G_{(i)}(t)\right) I_{\{\mathfrak{q}(r)K_{(i)}-\mathfrak{p}^{T}(r)G_{(i)}(t)>0\}} \mathrm{d}\left(-\bar{\xi}^{(i)}(t)\right)\right] \\ &= \mathbb{E}\left[\int_{0}^{T}\left(\mathfrak{q}(r) K - \mathfrak{p}^{\top}(r) G(t)\right) \mathrm{d}\left(\xi(t) - \bar{\xi}(t)\right) \mathrm{d}t\right] \\ &\geq 0. \end{aligned}$$

Thus

$$\mathbb{E}\left[\sum_{i=1}^{m}\int_{0}^{T}\left(\mathfrak{q}(r) K_{(i)} - \mathfrak{p}^{\top}(r) G_{(i)}(t)\right) I_{\{\mathfrak{q}(r)K_{(i)}-\mathfrak{p}^{T}(r)G_{(i)}(t)>0\}} \mathrm{d}\left(-\bar{\xi}^{(i)}(t)\right) \mathrm{d}t\right] = 0$$

which proves (17). Next we show that (16) is valid. For that, let us define the events:

$$\mathcal{A}^{(i)} \triangleq \{(t,\omega) \in [0,T] \times \Omega : \mathfrak{q}(r) K_{(i)} - \mathfrak{p}^{\top}(r) G_{(i)}(t) < 0\},$$

where $t \in [0,T]$, $1 \leq i \leq m$.

Define the stochastic process $\breve{\xi}^{(i)} : [0,T] \times \Omega \to [0,\infty)$ by

$$\breve{\xi}^{(i)}(t) = \bar{\xi}^{(i)}(t) + \int_{0}^{t} I_{\mathcal{A}^{(i)}}(r,\omega) \mathrm{d}r.$$

Then one can easily check that $\breve{\xi} = \left(\breve{\xi}^{(1)}, \breve{\xi}^{(2)}, \ldots, \breve{\xi}^{(m)}\right)$ is a measurable, adapted process which is nondecreasing left-continuous with right limits and $\breve{\xi}(0) = 0$, and which satisfies

$$P\left\{\sum_{i}\int_{0}^{T} G(s) \mathrm{d}\breve{\xi}^{(i)}(s) < \infty\right\}.$$

Further, we have

$$\begin{aligned} &\mathbb{E}\left[\int_{0}^{T}\left(\mathfrak{q}(r) K - \mathfrak{p}^{\top}(r) G(t)\right) \mathrm{d}\left(\breve{\xi}(t) - \bar{\xi}(t)\right) \mathrm{d}t\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{m}\int_{0}^{T}\left(\mathfrak{q}(r) K_{(i)} - \mathfrak{p}^{\top}(r) G_{(i)}(t)\right) I_{\mathcal{A}^{(i)}} \mathrm{d}t\right] \\ &< 0, \end{aligned}$$

which obviously contradicts to (22), unless for any i, we have $(Leb \otimes P)\left\{\mathcal{A}^{(i)}\right\} = 0$. We thus complete the proof. □

Remark 5. One can easily check that

$$\mathfrak{q}(s)=\exp\left\{\int_t^s\left[\bar{f}_y(r)-\frac{1}{2}\left|\bar{f}_z(r)\right|^2\right]\mathrm{d}r+\int_0^s\bar{f}_z(r)\,\mathrm{d}W(r)\right\},$$

which implies that $\mathfrak{q}(r)>0$, $r\in[0,T]$, P-a.s. So $-\mathfrak{p}(\cdot)/\mathfrak{q}(\cdot)$ makes sense. Clearly, our Theorem 3 for optimal singular control is completely different from [3]. Ours contains two variables $(\mathfrak{p}(\cdot),\mathfrak{q}(\cdot))$. As a matter of fact, we have

$$P\left\{K_{(i)}+\left(-\frac{\mathfrak{p}^\top(t)}{\mathfrak{q}(t)}\right)G_{(i)}(t)\geq 0,\ t\in[0,T],\ \forall i\right\}=1.$$

We claim that $-\mathfrak{p}(\cdot)/\mathfrak{q}(\cdot)$ is the partial derivative of value function, which will be studied in Section 4.3.

3.2. Optimal regular control

In this subsection, we study the optimal regular controls for systems driven by FBSDEs (8) under the types of Pontryagin, namely, necessary maximum principles for optimal control. To this end, we fix $\bar{\xi}\in\mathcal{U}_2$ and introduce the following convex perturbation control. Taking $u(\cdot)\in\mathcal{U}_1$, we define $v(\cdot)=u(\cdot)-\bar{u}(\cdot)$, $u^\varepsilon(\cdot)=\bar{u}(\cdot)+\varepsilon v(\cdot)$, where $\varepsilon>0$ is sufficiently small. Since U is convex, $u^\varepsilon(\cdot)\in\mathcal{U}(0,T)$. Let $(x^\varepsilon,y^\varepsilon,z^\varepsilon,u^\varepsilon)$ be the trajectory of the control system (8) corresponding to the control u^ε. Put $\delta x(\cdot)=x^\varepsilon(\cdot)-\bar{x}(\cdot)$. When $l=b$, σ and Φ, we denote

$$\begin{aligned}
l_x(t)&=l_x(t,\bar{x}(t),\bar{u}(t)),\ l_u(t)=l_u(t,\bar{x}(t),\bar{u}(t)),\\
l_{xx}(t)&=l_{xx}(t,\bar{x}(t),\bar{u}(t)),\ l_{xu}(t)=l_{xu}(t,\bar{x}(t),\bar{u}(t)),\\
l_{uu}(t)&=l_{uu}(t,\bar{x}(t),\bar{u}(t)),\ f_x(t)=f_x(t,\bar{x}(t),\bar{y}(t),\bar{z}(t),\bar{u}(t)),\\
f_y(t)&=f_y(t,\bar{x}(t),\bar{y}(t),\bar{z}(t),\bar{u}(t)),\ f_z(t)=f_z(t,\bar{x}(t),\bar{y}(t),\bar{z}(t),\bar{u}(t)),\\
f_{xx}(t)&=f_{xx}(t,\bar{x}(t),\bar{y}(t),\bar{z}(t),\bar{u}(t)),\ f_{xu}(t)=f_{xu}(t,\bar{x}(t),\bar{y}(t),\bar{z}(t),\bar{u}(t)),\\
f_{yy}(t)&=f_{yy}(t,\bar{x}(t),\bar{y}(t),\bar{z}(t),\bar{u}(t)),\cdots\ etc.
\end{aligned}$$

Let us introduce the following two kinds of variational equations, mainly taken from [15]. For simplicity, we omit the superscript.

$$\left\{\begin{array}{rcl}\mathrm{d}x_1(t)&=&[b_x(t)x_1(t)+b_u(t)v(t)]\,\mathrm{d}t+[\sigma_x(t)x_1(t)+\sigma_u(t)v(t)]\,\mathrm{d}W(t),\\ x_1(0)&=&0,\qquad t\in[0,T],\end{array}\right.\tag{23}$$

and

$$\left\{\begin{array}{rcl}\mathrm{d}x_2(t)&=&\Big[b_x(t)x_2(t)+x_1^\top(t)b_{xx}(t)x_1(t)\\ &&+2v(t)^\top b_{xu}(t)x_1(t)+v^\top(t)b_{uu}(t)v(t)\Big]\mathrm{d}t,\\ &&\Big[\sigma_x(t)x_2(t)+x_1^\top(t)\sigma_{xx}(t)x_1(t)\\ &&+2v(t)^\top\sigma_{xu}(t)x_1(t)+v^\top(t)\sigma_{uu}(t)v(t)\Big]\mathrm{d}W(t),\\ x_2(0)&=&0,\qquad t\in[0,T].\end{array}\right.\tag{24}$$

From Lemma 3.5 and Lemma 3.11 in [15], we have following result.

Lemma 6. *Assume that* (A1)-(A2) *is in force. Then, we have, for any* $\beta \geq 2$,

$$\|x_1\|_\infty^\beta \leq C,\ \|x_2\|_\infty^\beta \leq C,\ \|\delta x\|_\infty^\beta \leq C\varepsilon^\beta,$$
$$\|\delta x - \varepsilon x_1\|_\infty^\beta \leq C\varepsilon^{2\beta},\ \left\|\delta x - \varepsilon x_1 - \frac{\varepsilon^2}{2}x_2\right\|_\infty^\beta \leq C\varepsilon^{3\beta},$$

where

$$\|x_1\|_\infty^\beta = \left[\mathbb{E}\left(\sup_{t\in[0,T]}|x_1(t)|^\beta\right)\right]^{\frac{1}{\beta}}.$$

We shall introduce the so called variational equations for FBSDEs (8) beginning from the following two adjoint equations:

$$\begin{cases} -\mathrm{d}p(t) &= \Gamma(t)\,\mathrm{d}t - q(t)\,\mathrm{d}W(t), \\ p(T) &= \Phi_x(\bar{x}(T)), \end{cases} \tag{25}$$

and

$$\begin{cases} -\mathrm{d}P(t) &= \Pi(t)\,\mathrm{d}t - Q(t)\,\mathrm{d}W(t), \\ P(T) &= \Phi_{xx}(\bar{x}(T)), \end{cases} \tag{26}$$

where $\Gamma(\cdot)$, $\Pi(\cdot)$ are unknown two processes to be determined. Next we will derive two kinds of adjoint equations. The main idea is borrowed from [40]. First of all, we observe that

$$\begin{aligned}\Phi(x^\varepsilon(T)) - \Phi(\bar{x}(T)) &= \langle \delta x(T), \Phi_x(\bar{x}(T))\rangle \\ &\quad + \frac{1}{2}\langle \Phi_{xx}(\bar{x}(T))\,\delta x(T), \delta x(T)\rangle + o\left(\varepsilon^2\right) \\ &= \left\langle \varepsilon x_1(T) + \frac{1}{2}\varepsilon^2 x_2(T), \Phi_x(\bar{x}(T))\right\rangle \\ &\quad + \frac{1}{2}\varepsilon^2\langle \Phi_{xx}(\bar{x}(T))\,x_1(T), x_1(T)\rangle + o\left(\varepsilon^2\right),\end{aligned}$$

which inspires us to use the adjoint equations to expand the following:

$$\left\langle \varepsilon x_1(t) + \frac{1}{2}\varepsilon^2 x_2(t), p(t)\right\rangle + \frac{1}{2}\varepsilon^2 \mathrm{tr}\left[P(t)\,x_1(t)\,x_1^\top(t)\right] \text{ on } [0,T]. \tag{27}$$

Itô's formula applied to (27) yields for $t \in [0,T]$,

$$\begin{aligned}&\left\langle \varepsilon x_1(T) + \frac{1}{2}\varepsilon^2 x_2(T), p(T)\right\rangle + \frac{1}{2}\varepsilon^2 \mathrm{tr}\left[P(T)\,x_1(T)\,x_1^\top(T)\right] \\ &= \left\langle \varepsilon x_1(t) + \frac{1}{2}\varepsilon^2 x_2(t), p(t)\right\rangle + \frac{1}{2}\varepsilon^2 \mathrm{tr}\left[P(t)\,x_1(t)\,x_1^\top(t)\right] \\ &\quad + \int_t^T \left[\left\langle\left(\varepsilon x_1(s) + \frac{1}{2}\varepsilon^2 x_2(s)\right), \Lambda_1(s)\right\rangle + \varepsilon^2\langle x_1(s), \Lambda_2(s)\rangle \right. \\ &\quad \left. + \frac{1}{2}\varepsilon^2\langle \Lambda_3(s)\,x_1(s), x_1(s)\rangle + \Lambda_4(s)\right]\mathrm{d}s\end{aligned}$$

$$+\int_t^T \left[\left\langle \varepsilon x_1(s) + \frac{1}{2}\varepsilon^2 x_2(s), \Lambda_5(s) \right\rangle + \varepsilon^2 \langle x_1(s), \Lambda_6(s) \rangle \right.$$
$$\left. + \frac{1}{2}\varepsilon^2 \langle \Lambda_7(s) x_1(s), x_1(s) \rangle + \Lambda_8(s) \right] \mathrm{d}W(s),$$

where

$$\begin{aligned}
\Lambda_1(t) &= b_x^\top(t) p(t) + \sigma_x^\top(t) q(t) - \Gamma(t), \\
\Lambda_2(t) &= b_{xu}^\top(t) v(t) p(t) + \sigma_{xu}^\top(t) v(t) q(t) + P(t) b_u(t) v(t) \\
&\quad + \sigma_x^\top(t) P(t) \sigma_u(t) v(t) + Q(t) \sigma_u(t) v(t), \\
\Lambda_3(t) &= p(t) b_{xx}(t) + q(t) \sigma_{xx}(t) + P(t) b_x(t) + b_x^\top(t) P(t) \\
&\quad + \sigma_x^\top(t) P(t) \sigma_x(t) + Q(t) \sigma_x(t) + \sigma_x^\top(t) Q(t) - \Pi(t), \\
\Lambda_4(t) &= \varepsilon \langle p(t) b_u(t), v(t) \rangle + \frac{1}{2}\varepsilon^2 \langle p(t) b_{uu}(t) v(t), v(t) \rangle \\
&\quad + \varepsilon \langle q(t) \sigma_u(t), v(t) \rangle + \frac{1}{2}\varepsilon^2 \langle q(t) \sigma_{uu}(t) v(t), v(t) \rangle \\
&\quad + \frac{1}{2}\varepsilon^2 \left\langle \sigma_u(t) P(t) \sigma_u^\top v(t), v(t) \right\rangle, \\
\Lambda_5(t) &= p(t) \sigma_x(t) + q(t), \\
\Lambda_6(t) &= p(t) v(t) \sigma_{xu}(t) + P(t) \sigma_u(t) v(t), \\
\Lambda_7(t) &= p(t) \sigma_{xx}(t) + P(t) \sigma_x(t) + \sigma_x^\top(t) P(t) + Q(t), \\
\Lambda_8(t) &= \varepsilon \langle \sigma_u(t) p(t), v(t) \rangle + \frac{1}{2}\varepsilon^2 \langle p(t) \sigma_{uu}(t) v(t), v(t) \rangle.
\end{aligned}$$

Remark 7. Note that $\Gamma(t)$ and $\Pi(t)$ do not appear in the $\mathrm{d}W(s)$-term.

Define

$$\begin{cases} \mathrm{d}y^\varepsilon(t) &= -f((t, x^\varepsilon(t), y^\varepsilon(t), z^\varepsilon(t), u^\varepsilon(t)) \mathrm{d}t + z^\varepsilon(t) \mathrm{d}W(t), \\ y^\varepsilon(T) &= \Phi(x^\varepsilon(T)). \end{cases}$$

Let

$$\bar{y}^\varepsilon(t) = y^\varepsilon(t) - \left[\left\langle p(t), \left(\varepsilon x_1(t) + \frac{1}{2}\varepsilon^2 x_2(t) \right) \right\rangle + \frac{1}{2}\varepsilon^2 \langle P(t) x_1(t), x_1(t) \rangle \right] \tag{28}$$

$$\begin{aligned}\bar{z}^\varepsilon(t) = z^\varepsilon(t) - &\left[\left\langle \left(\varepsilon x_1(s) + \frac{1}{2}\varepsilon^2 x_2(s) \right), \Lambda_5(s) \right\rangle + \varepsilon^2 \langle x_1(s), \Lambda_6(s) \rangle \right. \\ &\left. + \frac{1}{2}\varepsilon^2 \langle \Lambda_7(s) x_1(s), x_1(s) \rangle + \Lambda_8(s) \right]. \end{aligned} \tag{29}$$

Clearly, from Lemma 6, we have

$$\Phi(x^\varepsilon(T)) = \Phi(\bar{x}(T)) + \left\langle p(T), \left(\varepsilon x_1(T) + \frac{1}{2}\varepsilon^2 x_2(T) \right) \right\rangle$$
$$+ \frac{1}{2}\varepsilon^2 \langle P(t) x_1(t), x_1(t) \rangle + o(\varepsilon^2).$$

After some tedious computations, we have

$$\begin{aligned}\overline{y}^{\varepsilon}(t) = \Phi(\bar{x}(T)) + \int_t^T \Bigg[& f(s, x^{\varepsilon}(s), y^{\varepsilon}(s), z^{\varepsilon}(s), u^{\varepsilon}(s)) \\ & + \left\langle \left(\varepsilon x_1(s) + \frac{1}{2}\varepsilon^2 x_2(s) \right), \Lambda_1(s) \right\rangle + \varepsilon^2 \langle x_1(s), \Lambda_2(s) \rangle \\ & + \frac{1}{2}\varepsilon^2 \langle \Lambda_3(s) x_1(s), x_1(s) \rangle + \Lambda_4(s) \Bigg] \mathrm{d}s - \int_t^T \overline{z}^{\varepsilon}(s) \,\mathrm{d}W(s) + o\left(\varepsilon^2\right).\end{aligned} \tag{30}$$

Put

$$\hat{y}^{\varepsilon}(t) = \overline{y}^{\varepsilon}(t) - \bar{y}(t), \ \ \hat{z}^{\varepsilon}(t) = \overline{z}^{\varepsilon}(t) - \bar{z}(t),$$

then we attain

$$\begin{aligned}\hat{y}^{\varepsilon}(t) = o\left(\varepsilon^2\right) + \int_t^T \Bigg[& f(s, x^{\varepsilon}(s), y^{\varepsilon}(s), z^{\varepsilon}(s), u^{\varepsilon}(s)) \\ & - f(s, \bar{x}(s), \bar{y}(s), \bar{z}(s), \bar{u}(s)) \\ & + \left\langle \left(\varepsilon x_1(s) + \frac{1}{2}\varepsilon^2 x_2(s) \right), \Lambda_1(s) \right\rangle + \varepsilon^2 \langle x_1(s), \Lambda_2(s) \rangle \\ & + \frac{1}{2}\varepsilon^2 \langle \Lambda_3(s) x_1(s), x_1(s) \rangle + \Lambda_4(s) \Bigg] \mathrm{d}s - \int_t^T \hat{z}^{\varepsilon}(s) \,\mathrm{d}W(s),\end{aligned} \tag{31}$$

where

$$\begin{aligned} & f(s, x^{\varepsilon}(s), y^{\varepsilon}(s), z^{\varepsilon}(s), u^{\varepsilon}(s)) - f(s, \bar{x}(s), \bar{y}(s), \bar{z}(s), \bar{u}(s)) \\ = & f\left(s, \bar{x}(s), \bar{y}(s), \bar{z}(s) + \varepsilon p(t) \sigma_u(t) v(t) + \frac{\varepsilon^2}{2} p(t) \sigma_{uu}(t) v^2(t), \bar{u}(s) \right) \\ & - f(s, \bar{x}(s), \bar{y}(s), \bar{z}(s), \bar{u}(s)) + f(s, x^{\varepsilon}(s), y^{\varepsilon}(s), z^{\varepsilon}(s), u^{\varepsilon}(s)) \\ & - f\left(s, \bar{x}(s), \bar{y}(s), \bar{z}(s) + \varepsilon p(t) \sigma_u(t) v(t) + \frac{\varepsilon^2}{2} p(t) \sigma_{uu}(t) v^2(t), \bar{u}(s) \right) \\ = & f\left(s, \bar{x}(s), \bar{y}(s), \bar{z}(s) + \varepsilon p(t) \sigma_u(t) v(t) + \frac{\varepsilon^2}{2} p(t) \sigma_{uu}(t) v^2(t), \bar{u}(s) \right) \\ & - f(s, \bar{x}(s), \bar{y}(s), \bar{z}(s), \bar{u}(s)) \\ & + f\Big(s, \bar{x}(s) + \varepsilon x_1(s) + \frac{\varepsilon^2}{2} x_2(s), \bar{y}(s) + \hat{y}^{\varepsilon}(t) + \Lambda_9(s), \\ & \bar{z}(s) + \hat{z}^{\varepsilon}(t) + \Lambda_{10}(s), u^{\varepsilon}(s) \Big) - f(s, \bar{x}(s), \bar{y}(s), \bar{z}(s), \bar{u}(s)) + o\left(\varepsilon^2\right).\end{aligned}$$

Next we are going to seek $\Gamma(\cdot)$, $\Pi(\cdot)$, determined by the optimal quadruple $(\bar{x}(\cdot), \bar{y}(\cdot), \bar{z}(\cdot), \bar{u}(\cdot))$, such that

$$
\begin{aligned}
& f\left(s, \bar{x}(s)+\varepsilon x_1(s)+\frac{1}{2}\varepsilon^2 x_2(s), \bar{y}(s)+\Lambda_9(s), \bar{z}(s)+\Lambda_{10}(s), u^{\varepsilon}(s)\right) \\
& -f(s, \bar{x}(s), \bar{y}(s), \bar{z}(s), \bar{u}(s))+\left\langle\left(\varepsilon x_1(s)+\frac{1}{2}\varepsilon^2 x_2(s)\right), \Lambda_1(s)\right\rangle \\
& +\frac{1}{2}\varepsilon^2\left\langle\Lambda_3(s) x_1(s), x_1(s)\right\rangle=o\left(\varepsilon^2\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\Lambda_9(s)= & \left\langle\left(\varepsilon x_1(s)+\frac{1}{2}\varepsilon^2 x_2(s)\right), p(s)\right\rangle+\frac{1}{2}\varepsilon^2\left\langle P(s) x_1(s), x_1(s)\right\rangle, \\
\Lambda_{10}(s)= & \left[\left\langle\left(\varepsilon x_1(s)+\frac{1}{2}\varepsilon^2 x_2(s)\right), \Lambda_5(s)\right\rangle+\varepsilon^2\left\langle x_1(s), \Lambda_6(s)\right\rangle\right. \\
& \left.+\frac{1}{2}\varepsilon^2\left\langle\Lambda_7(s) x_1(s), x_1(s)\right\rangle+\Lambda_8(s)\right],
\end{aligned}
$$

in which $o\left(\varepsilon^2\right)$ does not involve the terms $x_1(\cdot)$ and $x_2(\cdot)$. Note that in BSDE (30), there appears the term $x_1^{\top}(s)\Lambda_3(s) x_1(s)$. Hence, we make use of Taylor's expansion to

$$
\begin{aligned}
& f\left(s, \bar{x}(s)+\varepsilon x_1(s)+\frac{1}{2}\varepsilon^2 x_2(s), \bar{y}(s)+\Lambda_9(s), \bar{z}(s)+\Lambda_{10}(s), u^{\varepsilon}(s)\right) \\
= & f(s, \bar{x}(s), \bar{y}(s), \bar{z}(s), \bar{u}(s)) \\
& +\left[\varepsilon x_1(s)+\frac{1}{2}\varepsilon^2 x_2(s), \Lambda_9(s), \Lambda_{10}(s), \varepsilon v(s)\right] \cdot\left[f_x(s), f_y(s), f_z(s), f_u(s)\right]^{\top} \\
& +\frac{1}{2}\left[\varepsilon x_1(s)+\frac{1}{2}\varepsilon^2 x_2(s), \Lambda_9(s), \Lambda_{10}(s)\right] \cdot \mathbf{H}_1 f(s) \\
& \cdot\left[\varepsilon x_1(s)+\frac{\varepsilon^2}{2} x_2(s), \Lambda_9(s), \Lambda_{10}(s)\right]^{\top}+\varepsilon v(s) f_{yu}(s) \Lambda_9(s) \\
& +\varepsilon v(s) f_{xu}(s)\left(\varepsilon x_1(s)+\frac{1}{2}\varepsilon^2 x_2(s)\right)+\varepsilon v(s) f_{zu}\Lambda_{10}(s)+\frac{1}{2}\varepsilon^2 v^2(s) f_{uu}(s),
\end{aligned}
$$

where the Hessian matrix $\mathbf{H}_1$ is with respect to (x, y, z).

Then, we obtain

$$
\begin{aligned}
\Gamma(t)= & b_x^{\top}(t) p(t)+f_x(t)+f_y(t) p(t) \\
& +\sigma_x(t)^{\top} q(t)+f_z(t)\left[\sigma_x(t)^{\top}+q(t)\right], \\
\Pi(t)= & P(t) b_x(t)+b_x^{\top}(t) P(t)+f_y(t) P(t)+b_{xx}^{\top}(t) p(t) \\
& +\left[Q(t) \sigma_x(t)+\sigma_x(t)^{\top} Q(t)+\sigma_{xx}(t)^{\top} q(t)+\sigma_x(t)^{\top} P(t) \sigma_x(t)\right] \\
& +\left[f_z(t) P(t) \sigma_x(t)+f_z(t) \sigma_x(t)^{\top} P(t)+f_z(t) Q(t)+f_z(t) \sigma_{xx}(t)^{\top} p(t)\right] \\
& +\left[I_{n\times n}, p(t), \sigma_x(t)^{\top} p(t)+q(t)\right] \cdot \mathbf{H}_1 f(t) \\
& \cdot\left[I_{n\times n}, p(t), \sigma_x(t)^{\top} p(t)+q(t)\right]^{\top},
\end{aligned}
$$

where $I_{n\times n}$ denotes the identity matrix.

In the classical theory of optimal control for FBSDEs (cf. [73,74]), there generally appear two groups of the first-order adjoint equations, for instance $(\mathfrak{p}(\cdot), \mathfrak{q}(\cdot))$ in Eqs. (15). The following proposition will establish the relationship between them with $p(\cdot)$ from (25), which is very useful to study the connection between maximum principle and dynamic programming (see Theorem 40 below).

Proposition 8. *Suppose that Assumptions* (A1)-(A2) *are in force. Then we have*

$$p(s) = -\frac{\mathfrak{p}^\top(s)}{\mathfrak{q}(s)}, \ s \in [t,T], \ P\text{-}a.s.,$$

where $p(\cdot)$ *and* $(\mathfrak{p}(\cdot), \mathfrak{q}(\cdot))$ *are solutions to FBSDEs (25) and (15), respectively.*

The proof is just to apply the Itô's formula to $-\mathfrak{p}^T(s)/\mathfrak{q}(s)$, so we omit it.

We define the *classical* Hamiltonian function:

$$H(t,x,y,z,u,p,q) = \langle p, b(t,x,u)\rangle + \langle q, \sigma(t,x,u)\rangle + f(t,x,y,z,u),$$

where $(t,x,y,z,u,p,q) \in [0,T] \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^k \times \mathbb{R}^n \times \mathbb{R}^n$.

Then, we have

$$\begin{aligned}
\hat{y}^\varepsilon(t) = o\left(\varepsilon^2\right) + \int_t^T \Bigg[& f\Big(s, \bar{x}(s), \bar{y}(s), \bar{z}(s) \\
& + \varepsilon \langle p(s)\sigma_u(s), v(s)\rangle + \frac{\varepsilon^2}{2}\langle p(s)\sigma_{uu}(s)v(s), v(s)\rangle, \bar{u}(s)\Big) \\
& - f(s, \bar{x}(s), \bar{y}(s), \bar{z}(s), \bar{u}(s)) + f_y(s)\hat{y}^\varepsilon(s) + f_z(s)\hat{z}^\varepsilon(s) \\
& + \varepsilon^2 \Big\langle \big[p(s)b_{xu}(s) + q(s)\sigma_{xu}(s) + P(s)b_u(s) \\
& + P(s)\sigma_x(s)\sigma_u(s) + Q(s)\sigma_u(s)\big]x_1(s), v(s)\Big\rangle \\
& + \varepsilon \langle [p(s)b_u(s) + q(s)\sigma_u(s) + f_u(s)], v(s)\rangle \\
& + \frac{\varepsilon^2}{2}\left\langle \left[q(s)\sigma_{uu}(s) + p(s)b_{uu}(s) + \varepsilon^2 f_{uu}(s) + \sigma_u^\top(s)P(s)\sigma_u(s)\right]v(s), v(s)\right\rangle \\
& + \varepsilon v^\top(s) f_{xu}(s)\left[\varepsilon x_1(s) + \frac{\varepsilon^2}{2}x_2(s)\right] + \varepsilon v^\top(s) f_{yu}(s)\Lambda_9(s) \\
& + \varepsilon v^\top(s) f_{zu}(s)\Lambda_{10}(s)\Bigg]\mathrm{d}s - \int_t^T \hat{z}^\varepsilon(s)\,\mathrm{d}W(s).
\end{aligned}$$

Namely,

$$\begin{aligned}
\hat{y}^\varepsilon(t) = o\left(\varepsilon^2\right) + \int_t^T \Big\{ & f_y(s)\hat{y}^\varepsilon(s) + f_z(s)\hat{z}^\varepsilon(s) + \varepsilon^2 x_1^\top(s)\left[H_{xu}(s)\right. \\
& + P(s)\sigma_x(s)\sigma_u(s) + Q(s)\sigma_u(s) + p(t)f_{yu}(s) + P(s)b_u(s)\big]v(s) \\
& + \varepsilon\left[H_u(s) + f_z(s)p(s)\sigma_u(s)\right]v(s) \\
& + \frac{\varepsilon^2}{2}v^\top(s)\left[H_{uu}(s) + \sigma_u^2(s)P(s) + f_{zz}(s)p^2(s)\sigma_u^2(s) + f_z(s)p\sigma_{uu}\right]v(s) \\
& + \varepsilon^2 v^\top(s) f_{zu}(s)\Big[(p(t)\sigma_x(t) + q(t))x_1(s) + p(t)\sigma_u(s)v(s)\Big]\Big\}\mathrm{d}s
\end{aligned}$$

$$-\int_t^T \hat{z}^\varepsilon(s)\,\mathrm{d}W(s). \tag{32}$$

Remark 9. Note that FBSDEs (32) are somewhat different from (22) in Hu [40]. Specifically, the term $A_4 x_1(s) I_{E_\varepsilon}(s)$ disappears in (22) since

$$\mathbb{E}\left[\left(\int_0^T |A_4(s)\, x_1(s)\, I_{E_\varepsilon}(s)|\,\mathrm{d}s\right)^\beta\right] = o\left(\varepsilon^\beta\right) \text{ for } \beta \geq 2$$

in [40] by using spike variational approach. Nevertheless, the corresponding term in our paper is just $\varepsilon^2 x_1^\top(s)\,\Lambda_2(s)$. We will see a moment later that this term is needed to define an extensive "Hamiltonian function" as follows.

Define

$$\begin{aligned}\mathbb{H}(t,x,y,z,u,p,q,P,Q) \triangleq{}& H_{xu}(t,x,y,z,u,p,q) + b_u(t,x,u)\,P \\ &+ Q\sigma_u(t,x,u) + P\sigma_x(t,x,u)\,\sigma_u(t,x,u) \\ &+ f_{yu}(t,x,y,z,u)\,p + f_{zu}(t,x,y,z,u)\,[p\sigma_x(t,x,u)+q],\end{aligned}$$

where

$$(t,x,y,z,p,q,P,Q) \in [0,T]\times\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}.$$

We now give the adjoint equation for BSDE (32) as follows:

$$\begin{cases}\mathrm{d}\chi(t) &=& f_y(t)\,\chi(t)\,\mathrm{d}t + f_z(t)\,\chi(t)\,\mathrm{d}W(t), \\ \chi(0) &=& 1.\end{cases} \tag{33}$$

Lemma 10. *Under the Assumptions* (A1)-(A2), *SDE (33) admits a unique adapted strong solution* $\chi(t) \in \mathcal{S}^2(0,T;\mathbb{R})$. *Moreover, we have*

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |\chi(t)|^l\right] < \infty,\ \ l \geq 1, \tag{34}$$

$$\mathbb{E}\left[\sup_{0\leq s\leq T} |\chi(s)|^2\right]^4 < \infty. \tag{35}$$

Proof. The first inequality can be obtained from Theorem 6.16 of [77]. We deal with the second one. By Itô's formula, we have

$$\sup_{0\leq t\leq T} \chi^2(t) = 1 + \int_0^T \chi^2(t)\left(2\,|f_y(t)| + f_z^2(t)\right)\mathrm{d}t + \sup_{0\leq t\leq T}\int_0^t 2\chi^2(s)\, f_z(s)\,\mathrm{d}W(s).$$

It follows that

$$\left(\sup_{0\leq t\leq T}\chi^2(t)\right)^4 = C\left[1 + \left(\int_0^T \chi^2(t)\left(2\,|f_y(t)| + f_z^2(t)\right)\mathrm{d}t\right)^4\right.$$

$$+\left(\sup_{0\le t\le T}\int_0^t 2\chi^2(s)\,f_z(s)\,\mathrm{d}W(s)\right)^4\Bigg].$$

But by the B-D-G inequality, we get

$$\mathbb{E}\left(\sup_{0\le t\le T}\int_0^t 2\chi^2(s)\,f_z(s)\,\mathrm{d}W(s)\right)^4 \le C\mathbb{E}\left[\int_0^t \chi^8(s)\left(f_z^2(t)\right)^2\mathrm{d}s\right].$$

Thus

$$\begin{aligned}\mathbb{E}\left[\left(\sup_{0\le t\le T}\chi^2(t)\right)^4\right] &= C\mathbb{E}\left[1+\left(\int_0^T \chi^2(t)\,\mathrm{d}t\right)^4+\int_0^t \chi^8(s)\,\mathrm{d}s\right]\\ &\le C\mathbb{E}\left[1+\int_0^t \chi^8(s)\,\mathrm{d}s\right]\\ &\le C\mathbb{E}\left[1+T\sup_{0\le t\le T}|\chi(t)|^8\right]\\ &<\infty.\end{aligned}$$

The second estimation comes from the Hölder inequality. We complete the proof. □

Set

$$\begin{aligned}y^\varepsilon(t) &= \bar{y}(t)+\left[p^\top(t)\left(\varepsilon x_1(t)+\frac{\varepsilon^2}{2}x_2(t)\right)+\frac{\varepsilon^2}{2}x_1^\top(t)\,P(t)\,x_1(t)\right]+\hat{y}^\varepsilon(t),\\ z^\varepsilon(t) &= \bar{z}(t)+\Bigg[\Lambda_5(s)\left(\varepsilon x_1(s)+\frac{\varepsilon^2}{2}x_2(s)\right)+\varepsilon^2\Lambda_6(s)\,x_1(s)\\ &\quad+\frac{\varepsilon^2}{2}x_1^\top(t)\,\Lambda_7(s)\,x_1(s)+\Lambda_8(s)\Bigg]+\hat{z}^\varepsilon(t).\end{aligned}$$

We are able to give the variational equations as follows:

$$\begin{aligned}y_1^\varepsilon(t) &= \varepsilon p^\top(t)\,x_1(t),\\ z_1^\varepsilon(t) &= \varepsilon\left[x_1^\top(s)\,\Lambda_5(s)+p^\top(t)\,\sigma_u(t)\,v(t)\right]\end{aligned}$$

and

$$\begin{aligned}y_2^\varepsilon(t) &= \frac{\varepsilon^2}{2}\left[p^\top(t)\,x_2(t)+x_1^T(t)\,P(t)\,x_1(t)\right]+\hat{y}^\varepsilon(t),\\ z_2^\varepsilon(t) &= \frac{\varepsilon^2}{2}\Bigg[\Lambda_5(s)\,x_2(s)+2\Lambda_6(s)\,x_1(s)\\ &\quad+x_1^\top(s)\,\Lambda_7(s)\,x_1(s)+p^\top(t)\,v^\top(t)\,\sigma_{uu}(t)\,v(t)\Bigg]+\hat{z}^\varepsilon(t).\end{aligned}$$

Obviously, we have

$$y^{\varepsilon}(0)-\bar{y}(0)=\hat{y}^{\varepsilon}(0)\geq 0, \text{ since the definition of value function.}$$

Lemma 11. *Under the Assumptions* (A1)-(A2), *we have the following estimation*

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{y}^{\varepsilon}(t)\right|^{2}+\int_{0}^{T}\left|\hat{z}^{\varepsilon}(s)\right|^{2}\mathrm{d}s\right]=O\left(\varepsilon^{2}\right). \tag{36}$$

Proof. To prove (36), we consider (32) again. From assumptions (A1)-(A2), one can check that the adjoint equations (25) and (26) have a unique adapted strong solution, respectively. Furthermore, by classical approach, we are able to get the following estimates for $\beta\geq 2$,

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left(\left|p(t)\right|^{\beta}+\left|P(t)\right|^{\beta}\right)+\left(\int_{0}^{T}\left(\left|q(t)\right|^{2}+\left|Q(t)\right|^{2}\right)\mathrm{d}t\right)^{\frac{\beta}{2}}\right]<+\infty. \tag{37}$$

Applying Lemma 1 to (32), we get the desired result. Indeed, since there appears a term

$$\varepsilon\left[H_{u}(s)+f_{z}(s)\,p(s)\,\sigma_{u}(s)\right]v(s),$$

in BSDE (32), so we have the estimation with $O\left(\varepsilon^{2}\right)$. We complete the proof. $\square$

We shall derive a variational inequality which is crucial to establish the necessary condition for optimal control. Before this, we introduce the following the other type of singular control using the Hamiltonian function:

Definition 12 *(Singular control in the classical sense).* We call a control $\breve{u}(\cdot)\in\mathcal{U}(0,T)$ a singular control in the classical sense if $\breve{u}(\cdot)$ satisfies

$$\begin{cases}\text{i) } H_{u}\left(t,\breve{x}(t),\breve{y}(t),\breve{z}(t),\breve{u}(t),\breve{p}(t),\breve{q}(t)\right)\\ \qquad +f_{z}\left(t,\breve{x}(t),\breve{y}(t),\breve{z}(t)\right)\cdot\sigma_{u}^{\top}\left(t,\breve{x}(t),\breve{u}(t)\right)\cdot\breve{p}(t)=0, \text{ a.s., a.e., } t\in[0,T];\\ \text{ii) } H_{uu}\left(t,\breve{x}(t),\breve{y}(t),\breve{z}(t),\breve{u}(t),\breve{p}(t),\breve{q}(t)\right)\\ \qquad +\sigma_{u}\left(t,\breve{x}(t),\breve{u}(t)\right)\cdot\breve{P}(s)\cdot\sigma_{u}^{\top}\left(t,\breve{x}(t),\breve{u}(t)\right)\\ \qquad +f_{z}\left(t,\breve{x}(t),\breve{y}(t),\breve{z}(t)\right)\cdot\sigma_{uu}^{\top}\left(t,\breve{x}(t),\breve{u}(t)\right)\cdot\breve{p}(s)\\ \qquad +2f_{zu}\left(t,\breve{x}(t),\breve{y}(t),\breve{z}(t)\right)\cdot\breve{p}^{\top}(t)\cdot\sigma_{u}\left(t,\breve{x}(t),\breve{u}(t)\right)\\ \qquad +f_{zz}\left(t,\breve{x}(t),\breve{y}(t),\breve{z}(t)\right)\cdot\sigma_{u}^{\top}\left(t,\breve{x}(t),\breve{u}(t)\right)\breve{p}(t)\cdot\breve{p}^{\top}(t)\,\sigma_{u}\left(t,\breve{x}(t),\breve{u}(t)\right)=0,\\ \qquad \text{a.s., a.e., } t\in[0,T];\end{cases} \tag{38}$$

where $\left(\breve{x}(\cdot),\breve{y}(\cdot),\breve{z}(\cdot)\right)$ denotes the state trajectories driven by $\breve{u}(\cdot)$. Moreover, $\left(\breve{p}(\cdot),\breve{q}(\cdot)\right)$ and $\left(\breve{P}(\cdot),\breve{Q}(\cdot)\right)$ denote the adjoint processes given respectively by (25) and (26) with $\left(\bar{x}(t),\bar{y}(t),\bar{z}(t),\bar{u}(t)\right)$ replaced by $\left(\breve{x}(\cdot),\breve{y}(\cdot),\breve{z}(\cdot),\breve{u}(\cdot)\right)$. If this $\breve{u}(\cdot)$ is also optimal, then we call it a singular optimal control in the classical sense.

Remark 13. Hu [40] first considers the forward-backward stochastic control problem whenever the diffusion term $\sigma(t,x,u)$ depends on the control variable u with non-convex control domain. In order to establish the stochastic maximum principle, he introduces the $\mathcal{H}$-function of the following type:

$$\begin{aligned}\mathcal{H}(t,x,y,z,u,p,q,P)&\triangleq pb(t,x,u)+q\sigma(t,x,u)\\ &\quad+\frac{1}{2}\left(\sigma(t,x,u)-\sigma(t,\bar{x},\bar{u})\right)^{\top}P\left(\sigma(t,x,u)-\sigma(t,\bar{x},\bar{u})\right)\end{aligned}$$

$$+f\left(t,x,y,z+p\left(\sigma\left(t,x,u\right)-\sigma\left(t,\bar{x},\bar{u}\right)\right),u\right).$$

Note that this Hamiltonian function is slightly different from Peng 1990 [64]. The main difference of this variational equations with those in (Peng 1990) [64] appears in the term $p\left(t\right)\delta\sigma\left(t\right)I_{E_{\varepsilon}}\left(t\right)$ (the similar term $\varepsilon p\left(t\right)\sigma_{u}\left(t\right)v\left(t\right)+\frac{\varepsilon^{2}}{2}p\left(t\right)\sigma_{uu}\left(t\right)v^{2}\left(t\right)$ in our paper) in variational equation for BSDE and maximum principle for the definition of $p\left(t\right)$ in the variation of z, which is $O(\varepsilon)$ for any order expansion of f. So it is not helpful to use the second-order Taylor expansion for treating this term. The stochastic maximum principle (see [40]) says that if $\left(\breve{x}\left(t\right),\breve{y}\left(t\right),\breve{z}\left(t\right),\breve{u}\left(t\right)\right)$ is an optimal pair, then

$$\begin{aligned}&\mathcal{H}\left(t,\breve{x}\left(t\right),\breve{y}\left(t\right),\breve{z}\left(t\right),\breve{u}\left(t\right),\breve{p}\left(t\right),\breve{q}\left(t\right),\breve{P}\left(t\right)\right)\\&=\min_{u\in U}\mathcal{H}\left(t,\breve{x}\left(t\right),\breve{y}\left(t\right),\breve{z}\left(t\right),u,\breve{p}\left(t\right),\breve{q}\left(t\right),\breve{P}\left(t\right)\right).\end{aligned}\tag{39}$$

Apparently, Definition 12 says that a singular control in the classical sense is the real one that fulfills trivially the first and second-order necessary conditions in classical optimization theory dealing with the maximization problem (39), namely,

$$\begin{cases}\mathcal{H}_{u}\left(t,\breve{x}\left(t\right),\breve{y}\left(t\right),\breve{z}\left(t\right),\breve{u}\left(t\right),\breve{p}\left(t\right),\breve{q}\left(t\right),\breve{P}\left(t\right)\right)=0,\ \text{a.s., a.e., } t\in\left[0,T\right];\\\mathcal{H}_{uu}\left(t,\breve{x}\left(t\right),\breve{y}\left(t\right),\breve{z}\left(t\right),\breve{u}\left(t\right),\breve{p}\left(t\right),\breve{q}\left(t\right),\breve{P}\left(t\right)\right)=0,\ \text{a.s., a.e., } t\in\left[0,T\right].\end{cases}\tag{40}$$

It is easy to verify that (38) is equivalent to (40). Certainly, one could investigate stochastic singular optimal controls for forward-backward stochastic systems in other senses, say, in the sense of process in Skorohod space, which can be seen in Zhang [79] via viscosity solution approach (Hamilton-Jacobi-Bellman inequality), or in the sense of Pontryagin-type maximum principle (cf. Tang [71]). As this complete remake of the various topics is much longer than the present paper, it will be reported elsewhere.

Lemma 14 *(Variational inequality). Under the Assumptions* (A1)-(A2)*, it holds that*

$$\begin{aligned}0\leq\mathbb{E}\Bigg\{&\chi\left(T\right)o\left(\varepsilon^{2}\right)+\int_{0}^{T}\chi\left(s\right)\Big[\varepsilon\left\langle f_{z}\left(s\right)\sigma_{u}^{\top}\left(s\right)p\left(s\right)+H_{u}\left(s\right),v\left(s\right)\right\rangle\\&+\frac{1}{2}\varepsilon^{2}\Big(\left\langle f_{z}\left(s\right)\sigma_{uu}^{\top}\left(s\right)p\left(s\right)v\left(s\right),v\left(s\right)\right\rangle+2\left\langle\mathbb{H}\left(s\right)x_{1}\left(s\right),v\left(s\right)\right\rangle\\&+2\left\langle f_{zu}\left(s\right)p\left(t\right)\sigma_{u}\left(t\right)v\left(s\right),v\left(s\right)\right\rangle+\left\langle\tilde{\mathbb{H}}\left(s\right)v\left(s\right),v\left(s\right)\right\rangle\Big)\Big]\mathrm{d}s\Bigg\},\end{aligned}\tag{41}$$

where

$$\begin{aligned}\mathbb{H}\left(s\right)&=\mathbb{H}\left(t,\bar{x}\left(s\right),\bar{y}\left(s\right),\bar{z}\left(s\right),\bar{u}\left(s\right),p\left(s\right),q\left(s\right),P\left(s\right),Q\left(s\right)\right),\\\tilde{\mathbb{H}}\left(s\right)&=H_{uu}\left(s\right)+\sigma_{u}^{\top}\left(s\right)P\left(s\right)\sigma_{u}^{\top}\left(s\right)+f_{zz}\left(s\right)\sigma_{u}^{\top}\left(s\right)p^{\top}\left(s\right)p\left(s\right)\sigma_{u}\left(s\right).\end{aligned}$$

Proof. Using Itô's formula to $\left\langle\chi\left(s\right),\hat{y}^{\varepsilon}\left(s\right)\right\rangle$ on $[0,T]$, we get the desired result. □

Theorem 15. *Assume that* (A1)-(A2) *hold. If* $\bar{u}\left(\cdot\right)\in\mathcal{U}\left(0,T\right)$ *is a singular optimal control in the classical sense, then*

$$0\leq\mathbb{E}\left[\int_{0}^{T}\left\langle\chi\left(s\right)\mathbb{H}\left(s\right)x_{1}\left(s\right),v\left(s\right)\right\rangle\mathrm{d}s\right],\tag{42}$$

for any $v(\cdot) = u(\cdot) - \bar{u}(\cdot)$, $u(\cdot) \in \mathcal{U}(0,T)$.

Proof. According to the definition of value function, we have

$$\frac{J(u^\varepsilon) - J(\bar{u})}{\varepsilon^2} = \frac{y^\varepsilon(0) - \bar{y}(0)}{\varepsilon^2} = \frac{\hat{y}^\varepsilon(0)}{\varepsilon^2} \geq 0.$$

Letting $\varepsilon \to 0+$, we get the desired result from Definition 12 and Lemma 14. □

Remark 16. Clearly, if f does not depend on (y,z), then $\chi(\cdot) \equiv 1$. Consequently, (42) reduces to

$$\mathbb{E}\left[\int_0^T \langle \mathbb{H}(s) x_1(s), v(s)\rangle \mathrm{d}s\right] \geq 0,$$

which is just the classical case studied in Zhang et al. [80] for classical stochastic control problems. Meanwhile, our result actually extends Peng [66] to second order case.

Remark 17. Recall that, for deterministic system, it is possible to derive pointwise necessary conditions for optimal controls via the first suitable integral-type necessary conditions and normally there are no obstacles to establish the pointwise first-order necessary condition for optimal controls whenever an integral type one is on the hand. Nevertheless, the classical approach to handle the pointwise condition from the integral-type can not be employed directly in the framework of the pointwise second-order condition in the general stochastic setting because of certain feature the stochastic systems owning. In order to derive the second order variational equations for BSDE in Hu [40], the author there introduces two kinds of adjoint equations and a new Hamiltonian function. The main difference of this variational equations with those in (Peng 1990) [64] lies in the term $p(t)\delta\sigma(t) I_{E_\varepsilon}(t)$. Then, it is possible to get the maximum principle basing one variational equation. Note that the order of the difference between perturbed state, optimal state and first, second-order state is $o(\varepsilon)$.

As observed in Theorem 15, there appears a term $\langle \mathbb{H}(s) x_1(s), v(s)\rangle$. In order to deal with it, we give the expression of $x_1(\cdot)$, mainly taken from Theorem 1.6.14 in Yong and Zhou [77]. To this end, consider the following matrix-valued stochastic differential equation:

$$\left\{\begin{array}{rcl} \mathrm{d}\Psi(t) &=& b_x(t)\Psi(t)\mathrm{d}t + \sigma_x(t)\Psi(t)\mathrm{d}W(t), \\ \Psi(0) &=& I,\ t \in [0,T], \end{array}\right. \tag{43}$$

where I denotes the identity matrix in $\mathbb{R}^{n\times n}$. Then,

$$x_1(t) = \Psi(t)\left[\int_0^t \Psi^{-1}(r)(b_u(r) - \sigma_x(r)\sigma_u(r))v(r)\mathrm{d}r + \int_0^t \Phi^{-1}(r)\sigma_u(r)v(r)\mathrm{d}W(r)\right]. \tag{44}$$

Substituting the explicit representation (44) of x_1 into (42) yields

$$\begin{aligned} 0 \leq \mathbb{E}\Bigg[\int_0^T \chi(s)\Big\{ &\langle f_z(s) p(s) \sigma_{uu}(s) v(s), v(s)\rangle \\ &+\mathbb{H}(s) v(s) \Psi(s)\Big[\int_0^s \Psi^{-1}(r)(b_u(r) - \sigma_x(r)\sigma_u(r))v(r)\mathrm{d}r \end{aligned}$$

$$+\int_0^s \Psi^{-1}(r)\,\sigma_u(r)\,v(r)\,\mathrm{d}W(r)\Big] + \langle v(s), v(s) f_{zu}(s)\,p(t)\,\sigma_u(t)\rangle\Big\}\mathrm{d}s\Big]. \tag{45}$$

Clearly, (45) contains an Itô's integral. Next we shall borrow the spike variation method from [80] to check its order with perturbed control. More precisely, let $\varepsilon > 0$ and $E_\varepsilon \subset [0,T]$ be a Borel set with Borel measure $|E_\varepsilon| = \varepsilon$, define

$$u^\varepsilon(t) = \bar{u}(t)\,I_{E_\varepsilon^c}(t) + u(t)\,I_{E_\varepsilon}(t),$$

where $u(\cdot) \in \mathcal{U}(0,T)$. This u^ε is called a spike variation of the optimal control $\bar{u}$. For our aim, we only need to use $E_\varepsilon = [l, l+\varepsilon]$ for $l \in [0, T-\varepsilon]$ and $\varepsilon > 0$. Let

$$v(\cdot) = u^\varepsilon(\cdot) - \bar{u}(\cdot) = (u(t) - \bar{u}(\cdot))\,I_{E_\varepsilon}(\cdot).$$

Then, inserting it into (45), we have

$$\int_l^{l+\varepsilon} \chi(s)\,\mathbb{H}(s)\,(u^\varepsilon(s) - \bar{u}(s))\,\Psi(s)\int_l^s \Psi^{-1}(r)\,\sigma_u(r)\,(u^\varepsilon(r) - \bar{u}(r))\,\mathrm{d}W(r)\,\mathrm{d}s.$$

By Hölder inequality and Burkholder-Davis-Gundy inequality, we have

$$\begin{aligned}
&\mathbb{E}\left[\int_l^{l+\varepsilon} \chi(s)\,\mathbb{H}(s)\,v(s)\,\Psi(s)\int_0^s \Psi^{-1}(r)\,\sigma_u(r)\,v(r)\,\mathrm{d}W(r)\,\mathrm{d}s\right]\\
&\le \left[\mathbb{E}\int_l^{l+\varepsilon} |\chi(s)\,\mathbb{H}(s)\,v(s)\,\Psi(s)|^2\,\mathrm{d}s\right]^{\frac{1}{2}}\cdot\\
&\quad\left[\mathbb{E}\int_l^{l+\varepsilon}\int_l^s \left|\Psi^{-1}(r)\,\sigma_u(r)\,(u^\varepsilon(r) - \bar{u}(r))\right|^2\mathrm{d}r\mathrm{d}s\right]^{\frac{1}{2}}\\
&\le C\varepsilon^{\frac{3}{2}},
\end{aligned}$$

since $\sup_{s\in[0,T]} |\chi(s)|^2 < \infty$ from classical estimate for stochastic differential equations.

Lemma 18 *(Martingale representation theorem). Suppose that* $\phi \in L^2_{\mathbb{F}}\left(\Omega; L^2([0,T]:\mathbb{R}^n)\right)$*. Then, there exists a* $\kappa(\cdot,\cdot) \in L^2\left([0,T]; L^2_{\mathbb{F}}([0,T]\times\Omega;\mathbb{R}^n)\right)$ *such that*

$$\phi(t) = \mathbb{E}[\phi(t)] + \int_0^t \kappa(s,t)\,\mathrm{d}W(s), \ \text{a.s., a.e., } t \in [0,T].$$

The proof can be seen in Zhang et al. [80].

Lemma 19. *Assume that* (A1)-(A2) *hold. Then,*

$$\chi(\cdot)\,\mathbb{H}(\cdot) \in L^4_{\mathbb{F}}\left(\Omega; L^2([0,T]:\mathbb{R})\right).$$

Proof. We shall prove that

$$\mathbb{E}\left[\int_0^T |\chi(s)\,\mathbb{H}(t)|^2\,\mathrm{d}t\right]^2 < \infty.$$

From (A1)-(A2), we have

$$|\psi_x| \le C \text{ and } |\psi_{xu}| \le C \text{ for } \psi = b,\ \sigma,\ f.$$

Besides,

$$|f_{yu}| \le C, |f_{zu}| \le C.$$

Hence,

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T |\chi(t)\,\mathbb{H}(t)|^2\,\mathrm{d}t\right]^2 \le \mathbb{E}\left[\sup_{0\le t\le T}|\chi(t)|^2\int_0^T |\mathbb{H}(t)|^2\,\mathrm{d}t\right]^2\\
&\le \frac{1}{2}\mathbb{E}\left[\left(\sup_{0\le t\le T}|\chi(t)|^2\right)^4 + \left(\int_0^T |\mathbb{H}(t)|^2\,\mathrm{d}t\right)^4\right]\\
&\le \frac{1}{2}\mathbb{E}\left[\sup_{0\le t\le T}|\chi(t)|^2\right]^4 + \frac{1}{2}\mathbb{E}\left[\int_0^T \Big|H_{xu}(t,\bar{x}(t),\bar{y}(t),\bar{u}(t),\bar{z}(t),\bar{u}(t),\bar{p}(t),\bar{q}(t))\right.\\
&\quad +Q\sigma_u(t,\bar{x}(t),\bar{u}(t)) + P\sigma_x(t,\bar{x}(t),\bar{u}(t))\,\sigma_u(t,\bar{x}(t),\bar{u}(t))\\
&\quad +b_u(t,\bar{x}(t),\bar{u}(t))\,P(t) + f_{yu}(t,\bar{x}(t),\bar{y}(t),\bar{u}(t),\bar{z}(t),\bar{u}(t))\,p(t)\\
&\quad \left.+f_{zu}(t,\bar{x}(t),\bar{y}(t),\bar{u}(t),\bar{z}(t),\bar{u}(t))\,[p\sigma_x(t,\bar{x}(t),\bar{u}(t)) + q(t)]\Big|^2\mathrm{d}t\right]^4\\
&\le \frac{1}{2}\mathbb{E}\left[\sup_{0\le t\le T}|\chi(t)|^2\right]^4 + C + C\mathbb{E}\left[\int_0^T\left(|p(t)|^2+|q(t)|^2+|P(t)|^2+|Q(t)|^2\right)\mathrm{d}t\right]^4\\
&< \infty.
\end{aligned}$$

From Lemma 10 and the classical estimation in (37), we finish the proof. □

Therefore, by our assumption (A1)-(A2) and Lemma 18, for any $u \in U$, there exists a

$$\psi^u(\cdot,\cdot) \in L^2\left([0,T] : L^2_{\mathbb{F}}(\Omega\times[0,T]:\mathbb{R}^n)\right)$$

such that for a.e. $t \in [0,T]$

$$\chi(t)\,\mathbb{H}^\top(t)\,(u-\bar{u}(t)) = \mathbb{E}\left[\chi(t)\,\mathbb{H}^\top(t)\,(u-\bar{u}(t))\right] + \int_0^t \psi^u(s,t)\,dW(s). \tag{46}$$

Using (46), we are able to assert the following:

Theorem 20. *Suppose that* (A1)-(A2) *are in force. Let* $\bar{u}(\cdot)$ *be a singular optimal control in the classical sense, then we have*

$$\begin{aligned}&\mathbb{E}\left\langle \chi(r)\mathbb{H}(r)b_u(r)(u-\bar{u}(r)),u-\bar{u}(r)\right\rangle\\&+\partial_r^+\left(\chi(r)\mathbb{H}^\top(r)(u-\bar{u}(r)),\sigma_u(r)(u-\bar{u}(r))\right)\geq 0,\ a.e.,\ r\in[0,T],\end{aligned}$$

where

$$\begin{aligned}&\partial_r^+\left(\chi(r)\mathbb{H}^\top(r)(u-\bar{u}(r)),\sigma_u(r)(u-\bar{u}(r))\right)\\&=2\limsup_{\alpha\to 0}\frac{1}{\alpha^2}\mathbb{E}\int_r^{r+\alpha}\int_r^t\left\langle \psi^u(s,t),\Psi(r)\Psi^{-1}(s)\sigma_u(s)(u-\bar{u}(s))\right\rangle \mathrm{d}s\mathrm{d}t,\end{aligned}$$

where $\psi^u(s,t)$ *is obtained by (46), and* Ψ *is determined by (43).*

The proof is just to repeat the process in Theorem 3.10, [80], so we omit it.

Note that Theorem 20 is pointwise with respect to the time variable t (but also the integral form). Now if each of $\chi(\cdot)\mathbb{H}(\cdot)$ and $\bar{u}(\cdot)$ are regular enough, then the function $\psi^u(\cdot,\cdot)$ admits an explicit representation.

Suppose the following:

(A3) $\bar{u}(\cdot)\in\mathbb{L}_{2,\mathbb{F}}^{1,2}\left(\mathbb{R}^k\right)$, $\chi(\cdot)\mathbb{H}^\top(\cdot)\in\mathbb{L}_{2,\mathbb{F}}^{1,2}\left(\mathbb{R}^{k\times n}\right)\cap L^\infty\left([0,T]\times\Omega;\mathbb{R}^{k\times n}\right)$.

Theorem 21. *Suppose that the Assumptions* (A1)-(A3) *are in force. Let* $\bar{u}(\cdot)$ *be a singular optimal control in the classical sense, then we have*

$$\begin{aligned}&\left\langle \chi(r)\mathbb{H}(r)b_u(r)(u-\bar{u}(r)),u-\bar{u}(r)\right\rangle+\left\langle \nabla(\chi(r)\mathbb{H}(r))\sigma_u(r)(u-\bar{u}(r)),u-\bar{u}(r)\right\rangle\\&-\left\langle \chi(r)\mathbb{H}(r)\sigma_u(r)(u-\bar{u}(r)),\nabla\bar{u}(r)\right\rangle\geq 0,\ a.e.,\ r\in[0,T],\ \forall u\in U,\ P\text{-}a.s.\end{aligned}$$

Observe that the expression (42) is similar to (3.17) in [80]. Therefore, the proof is repeated as in Theorem 3.13 in Zhang and Zhang [80].

3.2.1. Example

We provide a concrete example to illustrate our theoretical result (Theorem 21) by looking at an example. If the FBSDEs considered in this paper are linear, it is possible to implement our principles directly. For convenience, we still adopt the notations introduced in Section 3.2.

Example 22. Consider the following FBSDEs with $n=1$ and $U=[-1,1]$.

$$\left\{\begin{array}{rcl}\mathrm{d}x(t)&=&\left[u(t)\sin x(t)+u^3(t)\right]\mathrm{d}t+x(t)u(t)\mathrm{d}W(t),\\-\mathrm{d}y(t)&=&\left[x^3(t)+x(t)u(t)+y(t)+z(t)+u^3(t)\right]\mathrm{d}t-z(t)\mathrm{d}W(t),\\x(0)&=&0,\ y(T)=x(T).\end{array}\right.\tag{47}$$

One can easily get the solutions to (33),

$$\chi(t)=\exp\left\{\frac{1}{2}t+W(t)\right\}>0,\ 0\leq t\leq T,\ P\text{-a.s.}$$

Set $(\bar{x}(t),\bar{y}(t),\bar{z}(t),\bar{u}(t))=(0,0,0,0)$. The corresponding adjoint equations are (25) and (26), namely,

$$\begin{cases} -\mathrm{d}p(t) &=& [p(t)+q(t)]\,\mathrm{d}t - q(t)\,\mathrm{d}W(t), \\ p(T) &=& 1, \end{cases} \tag{48}$$

and

$$\begin{cases} -\mathrm{d}P(t) &=& [P(t)+Q(t)]\,\mathrm{d}t - Q(t)\,\mathrm{d}W(t), \\ P(T) &=& 0. \end{cases} \tag{49}$$

We get immediately, the solutions to (48) and (49) are

$$\begin{cases} (p(t), q(t)) = \left(e^{T-t}, 0\right), \ t \in [0,T], \\ (P(t), Q(t)) = (0,0), \ t \in [0,T], \end{cases}$$

respectively. Hence, we have

$$\mathcal{H}_u(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t), p(t), q(t), P(t)) \equiv 0,$$
$$\mathcal{H}_{uu}(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t), p(t), q(t), P(t)) \equiv 0.$$

Therefore, $\bar{u}(t) = 0$ is a singular control in the classical sense. Moreover, we compute

$$\nabla\bar{u}(t) = 0, \ \mathbb{H}(t) = p(t)\cos\bar{x}(t) + 1 = e^{T-t} + 1.$$

Consequently, we get

$$\begin{aligned} &\langle \chi(r)\,\mathbb{H}(r)\,b_u(r)(u-\bar{u}(r)), u-\bar{u}(r)\rangle \\ &+ \langle \nabla(\chi(r)\,\mathbb{H}(r))\,\sigma_u(r)(u-\bar{u}(r)), u-\bar{u}(r)\rangle \\ &- \langle \chi(r)\,\mathbb{H}(r)\,\sigma_u(r)(u-\bar{u}(r)), \nabla\bar{u}(r)\rangle, \\ = {}& 0, \ \forall u \in [-1,1], \ \text{a.e. } r \in [0,T], \ P\text{-a.s.}, \end{aligned}$$

which indicates that Theorem 21 always holds and $\bar{u}(r) = 0$ is a singular optimal control.

4. Singular optimal controls via dynamic programming principle

In this section, we proceed our control problem from the view point of DPP. From now on, we focus on the following

$$\begin{cases} \mathrm{d}X^{t,x;u,\xi}(s) &=& b\left(s, X^{t,x;u,\xi}(s), u(s)\right)\mathrm{d}s + \sigma\left(s, X^{t,x;u,\xi}(s), u(s)\right)\mathrm{d}W(s) + G\mathrm{d}\xi(s), \\ \mathrm{d}Y^{t,x;u,\xi}(s) &=& -f\left(s, X^{t,x;u,\xi}(s), Y^{t,x;u,\xi}(s), Z^{t,x;u,\xi}(s), u(s)\right)\mathrm{d}s \\ && +Z^{t,x;u,\xi}(s)\,\mathrm{d}W(s) - K\mathrm{d}\xi(s), \\ X^{t,x;u,\xi}(t) &=& x, \ Y^{t,x;u,\xi}(T) = \Phi\left(X^{t,x;u,\xi}(T)\right), \qquad 0 \le t \le s \le T. \end{cases} \tag{50}$$

Since the value function defined by the solution of controlled BSDE (50), so from the existence and uniqueness, V defined in (11) is well-defined.

Remark 23. We assume that $G_{n\times m}$ and $K_{1\times m}$ are deterministic matrices. On the one hand, from the derivations in Theorem 5.1 of [39], it is convenient to show the “inaction” region for singular control; On the other hand, we may regard $Y^{t,x;u,\xi}(s) + K\xi(s)$ together as a solution, in this way, we are able to apply the classical Itô’s formula, avoiding the appearance of jump. We believe these assumptions can be removed properly, but at present, we consider constant only in our paper. Whilst in order to get the uniqueness of the solution to H-J-B inequality (51), we add the assumption $K^i > k_0 > 0$, $1 \le i \le m$. More details, see Theorem 2.2 in [79].

Set

$$\mathcal{L}(t,x,u)\varphi = \frac{1}{2}\mathrm{Tr}\left(\sigma\sigma^*(t,x,u)\varphi_{xx}\right) + \langle \varphi_x, b(t,x,u)\rangle,$$
$$(t,x,u) \in [0,T]\times\mathbb{R}^n\times U,\ \varphi \in C^{1,2}\left([0,T]\times\mathbb{R}^n\right).$$

4.1. Verification theorem via viscosity solutions

Zhang [79] has given a verification theorem for smooth solution of the following H-J-B inequality:

$$\begin{cases} \min\Big(V_x^\top(t,x)G + K, \frac{\partial}{\partial t}V(t,x) \\ +\min_{u\in U}\mathcal{L}(t,x,u)V(t,x) + f\left(t,x,V(t,x),V_x^\top(t,x)\sigma(t,x,u),u\right)\Big) = 0, \\ u(T,x) = \Phi(x),\ 0\le t\le T. \end{cases} \tag{51}$$

Lemma 24. *Define*

$$\mathcal{D}_t(V) := \left\{x\in\mathbb{R}^n : V(t,x) < V(t,x+Gh) + Kh,\ h\in\mathbb{R}_+^m,\ h\neq 0\right\}.$$

Then the optimal state process $X^{t,x;\hat{u},\hat{\xi}}$ is continuous whenever $\left(r, X_r^{t,x;\hat{u},\hat{\xi}}\right) \in \mathcal{D}_r(V)$. To be precise, we have

$$P\left(\Delta X_r^{t,x;\hat{u},\hat{\xi}} \neq 0,\ X_r^{t,x;\hat{u},\hat{\xi}} \in \mathcal{D}_r(V)\right) = 0,\ t\le r\le T.$$

The proof can be seen in Zhang [79].

Proposition 25. *Suppose that V is a classical solution of the H-J-B inequality (51) such that for some $l>1$,*

$$|V(t,x)| \le C\left(1+|x|^l\right).$$

Then for any $[0,T]\times\mathbb{R}^n$, $(u,\xi)\in\mathcal{U}$:

$$V(t,x) \le J(t,x,u,\xi).$$

Furthermore, if there exists $\left(\hat{u},\hat{\xi}\right)\in\mathcal{U}$ such that

$$1 = P\left\{\left(r, X^{t,x;\hat{u},\hat{\xi}}(r)\right)\in\mathcal{D}_r(V),\ 0\le r\le T\right\}, \tag{52}$$

$$1 = P\left\{\int_{[t,T]}\left[V_x^\top(r,x)G + K\right]\mathrm{d}r = 0\right\}, \tag{53}$$

$$1 = P\Bigg\{\left(s, X^{t,x;\hat{u},\hat{\xi}}(s+)\right)\in\mathcal{D}_r(V),\ t\le s\le T:$$
$$\hat{u}_s \in \min_{u\in U}\left[V_t\left(s, X^{t,x;\hat{u},\hat{\xi}}(s)\right) + \mathcal{L}\left(s, X^{t,x;\hat{u},\hat{\xi}}(s+),u\right)V\left(s, X^{t,x;\hat{u},\hat{\xi}}(s+)\right)\right.$$

$$+f\left(s,X^{t,x;\hat{u},\hat{\xi}}(s+),V\left(t,X^{t,x;\hat{u},\hat{\xi}}(s+)\right),\right.$$
$$\left.\left.\nabla V\left(s,X^{t,x;\hat{u},\hat{\xi}}(s+)\right)\sigma\left(s,X^{t,x;\hat{u},\hat{\xi}}(s+),u\right),u\right)\right]\Bigg\} \tag{54}$$

and

$$P\left\{V\left(s,X^{t,x;\hat{u},\hat{\xi}}(s)\right)=V\left(s,X^{t,x;\hat{u},\hat{\xi}}(s)\right)+K\Delta\hat{\xi}(s),\ t\leq s\leq T\right\}=1. \tag{55}$$

Then

$$V(t,x)=J\left(t,x;\hat{u}(\cdot),\hat{\xi}(\cdot)\right). \tag{56}$$

In this section, we remove the *unreal* condition, smooth on value function, by means of viscosity solutions.[3] We will recall the definition of a viscosity solution for H-J-B variational inequality (51) from [24]. Below, $\mathbb{S}^n$ will denote the set of $n\times n$ symmetric matrices.

Let us begin at introducing the following parabolic superjet:

Definition 26. Let $V(t,x)\in C([0,T]\times\mathbb{R}^n)$ and $(t,x)\in[0,T]\times\mathbb{R}^n$. We denote by $\mathcal{P}^{2,+}V(t,x)$, the "parabolic superjet" of V at (t,x) the set of triples $(p,q,X)\in\mathbb{R}\times\mathbb{R}^n\times\mathbb{S}^n$ which are such that

$$\begin{aligned}V(s,y)\leq{}&V(t,x)+p(s-t)+\langle q,x-y\rangle\\&+\frac{1}{2}\langle X(y-x),y-x\rangle+o\left(|s-t|+|y-x|^2\right).\end{aligned}$$

Similarly, we denote by $\mathcal{P}^{2,-}V(t,x)$, the "parabolic subjet" of V at (t,x) the set of triples $(p,q,X)\in\mathbb{R}\times\mathbb{R}^n\times\mathbb{S}^n$ which are such that

$$\begin{aligned}V(s,y)\geq{}&V(t,x)+p(s-t)+\langle q,x-y\rangle\\&+\frac{1}{2}\langle X(y-x),y-x\rangle+o\left(|s-t|+|y-x|^2\right).\end{aligned}$$

Lemma 27. *Let $V\in C([0,T]\times\mathbb{R}^n)$ and $(t,x)\in[0,T]\times\mathbb{R}^n$ be given. Then:*

1) $(p,q,X)\in\mathcal{P}^{2,+}V(t,x)$ if and only if there exists a function $\varphi\in C^{1,2}([0,T]\times\mathbb{R}^n)$ such that $V-\varphi$ attains a strict maximum at (t,x) and

$$(\varphi(t,x),\varphi_t(t,x),\varphi_x(t,x),\varphi_{xx}(t,x))=(V(t,x),p,q,X).$$

2) $(p,q,X)\in\mathcal{P}^{2,-}V(t,x)$ if and only if there exists a function $\varphi\in C^{1,2}([0,T]\times\mathbb{R}^n)$ such that $V-\varphi$ attains a strict minimum at (t,x) and

$$(\varphi(t,x),\varphi_t(t,x),\varphi_x(t,x),\varphi_{xx}(t,x))=(V(t,x),p,q,X).$$

More details can be seen in Lemma 5.4 and 5.5 in Yong and Zhou [77].

Define

$$\mathcal{G}(t,x,q,X)=\inf_{u\in U}\left\{\frac{1}{2}\mathrm{Tr}\left(\sigma\sigma^*(t,x,u)X\right)+\langle q,b(t,x,u)\rangle+f\left(t,x,V(t,x),q^\top\sigma(t,x,u),u\right)\right\}. \tag{57}$$

[3] In the classical optimal stochastic control theory, the value function is a solution to the corresponding H-J-B equation whenever it has sufficient regularity (Fleming and Rishel [31], Krylov [53]). Nevertheless, when it is only known that the value function is continuous, then, the value function is a solution to the H-J-B equation in the viscosity sense (see Lions [24]).

Definition 28. (i) It can be said $V(t,x) \in C([0,T] \times \mathbb{R}^n)$ is a viscosity subsolution of (51) if $V(T,x) \geq \Phi(x)$, $x \in \mathbb{R}^n$, and at any point $(t,x) \in [0,T] \times \mathbb{R}^n$, for any $(p,q,X) \in \mathcal{P}^{2,+}V(t,x)$,

$$\min(qG + K, p + \mathcal{G}(t,x,q,X)) \geq 0. \tag{58}$$

In other words, at any point (t,x), we have both $qG + K \geq 0$ and

$$p + \mathcal{G}(t,x,q,X) \geq 0.$$

(ii) It can be said $V(t,x) \in C([0,T] \times \mathbb{R}^n)$ is a viscosity supersolution of (51) if $V(T,x) \leq \Phi(x)$, $x \in \mathbb{R}^n$, and at any point $(t,x) \in [0,T] \times \mathbb{R}^n$, for any $(p,q,X) \in \mathcal{P}^{2,-}V(t,x)$,

$$\min(qG + K, p + \mathcal{G}(t,x,q,X)) \leq 0. \tag{59}$$

In other words, at any point where $qG + K \geq 0$, we have

$$p + \mathcal{G}(t,x,q,X) \leq 0.$$

(iii) It can be said $V(t,x) \in C([0,T] \times \mathbb{R}^n)$ is a viscosity solution of (51) if it is both a viscosity sub and super solution.

We have the following result:

Proposition 29. *Assume that* (A1)-(A2) *are in force. Then there exists at most one viscosity solution of H-J-B inequality* (51) *in the class of bounded and continuous functions.*

We need a generalized Itô's formula. Define

$$\begin{gathered}\mathcal{L}(t,x,u)\Psi = \frac{1}{2}\mathrm{Tr}\left(\sigma\sigma^*(t,x,u)D^2\Psi\right) + \langle D\Psi, b(t,x,u)\rangle, \\ (t,x,u) \in [0,T] \times \mathbb{R}^n \times U,\ \Psi \in C^{1,2}([0,T] \times \mathbb{R}^n).\end{gathered}$$

For any $\Psi \in C^{1,2}([0,T] \times \mathbb{R}^n; \mathbb{R})$, by virtue of Doléans-Dade-Meyer formula (see [20,39]), we have

$$\begin{aligned}\Psi(s,X_s) = \Psi(t,x) &+ \int_t^s \Psi_t(r,X(r)) + \mathcal{L}(r,X(r),v)\Psi(r,X(r))\,\mathrm{d}r \\ &+ \int_t^s \Psi_x(r,X(r))\sigma(r,X(r),v(r))\,\mathrm{d}W(r) + \int_t^s \Psi_x(r,X(r))G\mathrm{d}\xi(r) \\ &+ \sum_{t \leq r \leq s}\{\Psi(r,X(r+)) - \Psi(r,X(r)) - \Psi_x(r,X(r))\Delta X(r)\}.\end{aligned} \tag{60}$$

We begin to introduce a useful lemma.

Lemma 30. *Assume that* (A1)-(A2) *are in force. Let* $(t,x) \in [0,T) \times \mathbb{R}^n$ *be fixed and let* $\left(X^{t,x;u,\xi}(\cdot), u(\cdot), \xi(\cdot)\right)$ *be an admissible pair. Define processes*

$$\begin{cases} z_1(r) \doteq b\left(r, X^{t,x;u,\xi}(r), u(r)\right), \\ z_2(r) \doteq \sigma\left(r, X^{t,x;u,\xi}(r), u(r)\right)\sigma^*\left(r, X^{t,x;u,\xi}(r), u(r)\right), \\ z_3(r) \doteq f\left(r, X^{t,x;u,\xi}(r), Y^{t,x;u,\xi}(r), Z^{t,x;u,\xi}(r), u(r)\right). \end{cases}$$

Then

$$\lim_{h\to 0+}\frac{1}{h}\int_t^{t+h}\left|z_i(r)-z_i(t)\right|\mathrm{d}r=0,\quad a.e.\ t\in[0,T],\ i=1,2,3. \tag{61}$$

The proof can be found in [77] by similar arguments.

Lemma 31. *Let $g\in C([0,T])$. Extend g to $(-\infty,+\infty)$ with $g(t)=g(T)$ for $t>T$, and $g(t)=g(0)$, for $t<0$. Suppose that there is an integrable function $\rho\in L^1([0,T];\mathbb{R})$ and some $h_0>0$, such that*

$$\frac{g(t+h)-g(t)}{h}\leq\rho(t),\quad a.e.\ t\in[0,T],\qquad h\leq h_0.$$

Then

$$g(\beta)-g(\alpha)\leq\int_\alpha^\beta\limsup_{h\to 0+}\frac{g(t+h)-g(t)}{h}\mathrm{d}r,\ \forall 0\leq\alpha\leq\beta\leq T.$$

The proof can be seen in Zhang [78].

The main result in this section is the following.

Theorem 32 *(Verification theorem). Suppose that the Assumptions* (A1)-(A2) *are in force. Let $V\in C([0,T]\times\mathbb{R}^n)$, be a viscosity solution of the H-J-B equations (51), satisfying the following conditions:*

$$\begin{cases} i)\ V(t+h,x)-V(t,x)\leq C\left(1+|x|^m\right)h, \qquad m\geq 0,\\ \quad \text{for all } x\in\mathbb{R}^n, 0<t<t+h<T.\\ ii)\ V \text{ is semiconcave, uniformly in } t, \text{ i.e. there exists } C_0\geq 0\\ \quad \text{such that for every } t\in[0,T],\ V(t,\cdot)-C_0|\cdot|^2 \text{ is concave on } \mathbb{R}^n.\end{cases} \tag{62}$$

Then we have

$$V(t,x)\leq J(t,x;u(\cdot),\xi(\cdot)), \tag{63}$$

for any $(t,x)\in(0,T]\times\mathbb{R}^n$ and any $u(\cdot)\times\xi(\cdot)\in\mathcal{U}(t,T)$.

Furthermore, let $(t,x)\in[0,T)\times\mathbb{R}^n$ be fixed and let

$$\left(\bar{X}^{t,x;\bar{u},\bar{\xi}}(\cdot),\bar{Y}^{t,x;\bar{u},\bar{\xi}}(\cdot),\bar{Z}^{t,x;\bar{u},\bar{\xi}}(\cdot)\,\bar{u}(\cdot),\bar{\xi}(\cdot)\right)$$

be an admissible pair such that there exist a function $\varphi\in C^{1,2}([0,T];\mathbb{R}^n)$ and a triple

$$\left(\overline{p},\overline{q},\overline{\Theta}\right)\in\left(L^2_{\mathcal{F}_t}([t,T];\mathbb{R})\times L^2_{\mathcal{F}_t}([t,T];\mathbb{R}^n)\times L^2_{\mathcal{F}_t}([t,T];\mathbb{S}^n)\right) \tag{64}$$

satisfying

$$\begin{cases}\left(\bar{p}(s),\bar{q}(s),\bar{\Theta}(s)\right)\in\mathcal{P}^{2,+}V\left(s,\bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right),\\ \left(\frac{\partial\varphi}{\partial t}\left(s,\bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right),D_x\varphi\left(s,\bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right),D^2\varphi\left(s,\bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right)\right)=\left(\overline{p}(s),\overline{q}(s),\overline{\Theta}(s)\right),\\ \bar{p}(t)G+K=0,\ a.e.\ t\in[0,T],\ P\text{-}a.s.\end{cases} \tag{65}$$

and

$$\mathbb{E}\left[\int_t^T \left[\overline{p}(s) + \mathcal{G}\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s), \overline{\varphi}(s), \overline{p}(s), \overline{\Theta}(s), \overline{u}(s)\right)\right] \mathrm{d}s\right] \leq 0, \tag{66}$$

where $\overline{\varphi}(s) = \varphi\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right)$ *and* $\mathcal{G}$ *is defined in (57). Then* $(\bar{X}^{t,x;\bar{u},\bar{\xi}}(\cdot), \bar{u}(\cdot), \bar{\xi}(\cdot))$ *is an optimal pair.*

In order to prove Theorem 32, we need the following lemma:

Lemma 33. *Let* V *be a viscosity subsolution of the H-J-B equations (51) satisfying (62). Then we have*

$$\mathbb{E}\frac{1}{h}\left[V\left(s+h, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s+h)\right) - V\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right)\right] \leq \rho(s), \ 0 < h \leq T-t, \tag{67}$$

where $\rho(s) \in L^1([t,T]:\mathbb{R})$.

The proof can be seen in the Appendix.

Proof of Theorem 32. We have (63) from the uniqueness of viscosity solutions of the H-J-B equations (51). It remains to show that $\left(\bar{X}^{t,x;\bar{u},\bar{\xi}}(\cdot), \bar{u}(\cdot), \bar{\xi}(\cdot)\right)$ is an optimal, we now fix $t_0 \in [t,T]$ such that (64) and (65) hold at t_0. For $z_1(\cdot) = \overline{b}(\cdot)$, $z_2(\cdot) = \overline{\sigma}(\cdot)\overline{\sigma}(\cdot)^*$, $z_3(\cdot) = \overline{f}(\cdot)$. We claim that the set of such points is of full measure in $[t,T]$ by Lemma 7 in [78]. Now we fix $\omega_0 \in \Omega$ such that the regular conditional probability $P\left(\cdot|\mathcal{F}_{t_0}^t\right)(\omega_0)$, given $\mathcal{F}_{t_0}^t$ is well defined. In this new probability space, the random variables $\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0), \overline{p}(t_0), \overline{q}(t_0), \overline{\Theta}(t_0)$ are almost surely deterministic constants and equal to

$$\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0,\omega_0), \overline{p}(t_0,\omega_0), \overline{q}(t_0,\omega_0), \overline{\Theta}(t_0,\omega_0),$$

respectively. We remark that in this probability space the Brownian motion W is still a standard Brownian motion although now $W(t_0) = W(t_0,\omega_0)$ almost surely. The space is now equipped with a new filtration $\{\mathcal{F}_r^t\}_{t\leq r\leq T}$ and the control process $\overline{u}(\cdot)$ is adapted to this new filtration. For P-a.s. ω_0 the process $\bar{X}^{t,x;\bar{u},\bar{\xi}}(\cdot)$ is a solution of (1.1) on $[t_0,T]$ in $\left(\Omega,\mathcal{F},P\left(\cdot|\mathcal{F}_{t_0}^t\right)(\omega_0)\right)$ with the initial condition $\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0) = \bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0,\omega_0)$.

Then on the probability space $\left(\Omega,\mathcal{F},P\left(\cdot|\mathcal{F}_{t_0}^t\right)(\omega_0)\right)$, we are going to apply Itô's formula to φ on $[t_0,t_0+h]$ for any $h>0$,

$$\begin{aligned}
&\varphi\left(t_0+h, \bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0+h)\right) - \varphi\left(t_0, \bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0)\right)\\
=&\int_{t_0}^{t_0+h}\left[\frac{\partial\varphi}{\partial t}\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right) + \left\langle D_x\varphi\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right), \overline{b}(r)\right\rangle\right.\\
&\left.+\frac{1}{2}\mathrm{tr}\left\{\overline{\sigma}(r)^* D_{xx}\varphi\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)\overline{\sigma}(r)\right\}\right]\mathrm{d}r\\
&+\int_{t_0}^{t_0+h}\varphi_x\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)G\mathrm{d}\bar{\xi}(r) + \int_{t_0}^{t_0+h}\left\langle D_x\varphi\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right), \overline{\sigma}(r)\right\rangle \mathrm{d}W(r)\\
&+\sum_{t_0\leq r\leq t_0+h}\left\{\varphi\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right) - \varphi\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)\right.\\
&\left.-\varphi_x\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)\Delta\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right\}.
\end{aligned}$$

Taking conditional expectation value $\mathbb{E}^{\mathcal{F}_{t_0}^t}(\cdot)(\omega_0)$, dividing both sides by h, and using (65), we have

$$
\begin{aligned}
&\mathbb{E}\left[V\left(s+\theta,\bar{X}^{t,x;\bar{u},\bar{\xi}}(s+\theta)\right)-V\left(s,\bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right)\right]\\
&\geq \mathbb{E}\left[\varphi\left(s+\theta,\bar{X}^{t,x;\bar{u},\bar{\xi}}(s+\theta)\right)-\varphi\left(s,\bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right)\right]\\
&=\mathbb{E}\Bigg[\int_s^{s+\theta}\Big[\frac{\partial\varphi}{\partial t}\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)+\left\langle D_x\varphi\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right),\bar{b}(r)\right\rangle\\
&\quad+\frac{1}{2}\mathrm{tr}\left\{\overline{\sigma}(r)^{*}D_{xx}\varphi\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)\overline{\sigma}(r)\right\}\Big]\mathrm{d}r\\
&\quad+\int_{t_0}^{t_0+h}\varphi_x\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)G\mathrm{d}\bar{\xi}(r)+\sum_{s\leq r\leq s+\theta}\Big\{\varphi\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)\\
&\quad-\varphi\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)-\varphi_x\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)\Delta\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\Big\}\Bigg].
\end{aligned}
\tag{68}
$$

We now handle the last two terms. Note that

$$
\Delta\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)=G\Delta\xi(r)
$$

and

$$
\bar{X}^{t,x;\bar{u},\bar{\xi}}(r+)=\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)+\Delta\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)=\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)+G\Delta\xi(r).
$$

Thus

$$
\begin{aligned}
&-\mathbb{E}\left[\int_{t_0}^{t_0+h}\varphi_x\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)G\mathrm{d}\bar{\xi}(r)\right]+\mathbb{E}\left[\varphi_x\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)\Delta\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right]\\
&=-\mathbb{E}\left[\int_{t_0}^{t_0+h}\varphi_x\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)G\mathrm{d}\bar{\xi}^{c}(r)\right]\\
&=\mathbb{E}\left[\int_{t_0}^{t_0+h}K\mathrm{d}\bar{\xi}^{c}(r)\right].
\end{aligned}
\tag{69}
$$

We now deal the term

$$
\begin{aligned}
&-\mathbb{E}\left[\sum_{t_0\leq r\leq t_0+h}\left\{\varphi\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r+)\right)-\varphi\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)\right\}\right]\\
&=-\mathbb{E}\left[\sum_{t_0\leq r\leq t_0+h}\left\{\int_0^1\varphi_x\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)+\theta\Delta\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)G\Delta\bar{\xi}(r)\,\mathrm{d}\theta\right\}\right]\\
&=\mathbb{E}\left[K\Delta\bar{\xi}(r)\right].
\end{aligned}
\tag{70}
$$

Combining (69) and (70), we have

$$
\begin{aligned}
&\frac{1}{h}\mathbb{E}^{\mathcal{F}_{t_0}^t(\omega_0)}\left[V\left(t_0+h,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0+h)\right)-V\left(t_0,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0)\right)\right]\\
&=\frac{1}{h}\mathbb{E}^{\mathcal{F}_{t_0}^t(\omega_0)}\left\{\int_{t_0}^{t_0+h}\left[\frac{\partial\varphi}{\partial t}\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)+\left\langle D_x\varphi\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right),\bar{b}(r)\right\rangle\right.\right.\\
&\quad\left.\left.+\frac{1}{2}\mathrm{tr}\left\{\overline{\sigma}(r)^* D_{xx}\varphi\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)\overline{\sigma}(r)\right\}\right]\mathrm{d}r-\int_{t_0}^{t_0+h}K\mathrm{d}\bar{\xi}(r)\right\}.
\end{aligned}
\tag{71}
$$

Letting $h\to 0$, and employing the similar delicate method as in the proof of Theorem 4.1 of Gozzi et al. [37], we have

$$
\begin{aligned}
&\frac{1}{h}\limsup_{h\to 0+}\mathbb{E}^{\mathcal{F}_{t_0}^t(\omega_0)}\left[V\left(t_0+h,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0+h)\right)-V\left(t_0,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0)\right)\right]\\
&\leq\frac{\partial\varphi}{\partial t}\left(t_0,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0,\omega_0)\right)+\left\langle D_x\varphi\left(t_0,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0,\omega_0)\right),\bar{b}(t_0)\right\rangle\\
&\quad+\frac{1}{2}\mathrm{tr}\left\{\overline{\sigma}(t_0)^* D_{xx}\varphi\left(t_0,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0,\omega_0)\right)\overline{\sigma}(t_0)\right\}-K\mathrm{d}\bar{\xi}(t_0)\\
&=\overline{p}(t_0,\omega_0)+\left\langle\overline{q}(t_0,\omega_0),\bar{b}(t_0)\right\rangle+\frac{1}{2}\mathrm{tr}\left\{\overline{\sigma}(t_0)^*\overline{\Theta}(t_0,\omega_0)\overline{\sigma}(t_0)\right\}-K\mathrm{d}\bar{\xi}(t_0).
\end{aligned}
$$

From Lemma 33, that there exist $\rho\in L^1(t_0,T;\mathbb{R})$ and $\rho_1\in L^1(\Omega;\mathbb{R})$ such that

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{h}\left[V\left(t+h,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t+h)\right)-V\left(t,\bar{X}^{s,y;u,\xi}(t)\right)\right]\right]\\
&\leq\rho(t),\ \text{for } h\leq h_0,\ \text{for some } h_0>0
\end{aligned}
\tag{72}
$$

and

$$
\begin{aligned}
&\mathbb{E}^{\mathcal{F}_{t_0}^t(\omega_0)}\left[\frac{1}{h}\left[V\left(t+h,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t+h)\right)-V\left(t,\bar{X}^{s,y;u,\xi}(t)\right)\right]\right]\\
&\leq\rho_1(\omega_0),\ h\leq h_0,\ \text{for some } h_0>0.
\end{aligned}
\tag{73}
$$

holds, respectively.

By virtue of Fatou's Lemma, noting (73), we obtain

$$
\begin{aligned}
&\limsup_{h\to 0+}\frac{1}{h}\mathbb{E}\left[V\left(t_0+h,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0+h)\right)-V\left(t_0,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0)\right)\right]\\
&=\limsup_{h\to 0+}\frac{1}{h}\mathbb{E}\left[\mathbb{E}^{\mathcal{F}_{t_0}^t(\omega_0)}\left\{V\left(t_0+h,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0+h)\right)-V\left(t_0,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0)\right)\right\}\right]\\
&\leq\mathbb{E}\left[\limsup_{h\to 0+}\frac{1}{h}\mathbb{E}^{\mathcal{F}_{t_0}^t(\omega_0)}\left\{V\left(t_0+h,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0+h)\right)-V\left(t_0,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t_0)\right)\right\}\right]\\
&\leq\mathbb{E}\left[\overline{p}(t_0)+\left\langle\overline{q}(t_0),\bar{b}(t_0)\right\rangle+\frac{1}{2}\mathrm{tr}\left\{\overline{\sigma}(t_0)^*\overline{\Theta}(t_0)\overline{\sigma}(t_0)\right\}-K\mathrm{d}\bar{\xi}(t_0)\right],
\end{aligned}
\tag{74}
$$

for a.e. $t_0\in[t,T]$. Then the rest of the proof goes exactly as in [37].

We apply Lemma 8 in [78] to $g(t)=\mathbb{E}\left[V\left(t,\bar{X}^{t,x;\bar{u},\bar{\xi}}(t)\right)\right]$, using (72), (66) and (74) to get

$$
\begin{aligned}
&\mathbb{E}\left[V\left(T,\bar{X}^{t,x;\bar{u},\bar{\xi}}(T)\right)-V(t,x)\right]\\
&\leq \mathbb{E}\left[\int_t^T \overline{p}(t)+\left\langle \overline{q}(t),\overline{b}(t)\right\rangle+\frac{1}{2}\mathrm{tr}\left[\overline{\sigma}(t)^*\,\overline{\Theta}(t)\,\overline{\sigma}(t)\right]\mathrm{d}t-K\mathrm{d}\bar{\xi}(t_0)\right]\\
&\leq -\mathbb{E}\left[\int_t^T \overline{f}(t)\,\mathrm{d}t+K\mathrm{d}\bar{\xi}(t_0)\right].
\end{aligned}
$$

From this we claim that

$$
\begin{aligned}
V(t,x) &\geq \mathbb{E}\left[V\left(T,\bar{X}^{t,x;\bar{u},\bar{\xi}}(T)\right)+\int_t^T \overline{f}(r)\,\mathrm{d}r+\int_s^T K\mathrm{d}\bar{\xi}(r)\right]\\
&= \mathbb{E}\left[\Phi\left(\bar{X}^{t,x;\bar{u},\bar{\xi}}(T)\right)+\int_t^T \overline{f}(r)\,\mathrm{d}r+\int_s^T K\mathrm{d}\bar{\xi}(r)\right],
\end{aligned}
$$

where

$$
\bar{f}(r)=f\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r),V\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right),\bar{q}(r)\,\sigma\left(r,\bar{X}^{t,x;\bar{u},\bar{\xi}}(r),\bar{u}\right)\right).
$$

Thus, combining the above with the first assertion (63), we prove the $\left(\bar{X}^{t,x;\bar{u},\bar{\xi}}(\cdot),\overline{u}(\cdot)\right)$ is an optimal pair. The proof is thus completed. □

Remark 34. The condition (66) is just equivalent to the following:

$$
\begin{aligned}
\overline{p}(s) &= \min_{u\in U}\mathcal{G}\left(s,\bar{X}^{t,x;\bar{u},\bar{\xi}}(s),\overline{\varphi}(s),\overline{q}(s),\overline{\Theta}(s),u\right)\\
&= \mathcal{G}\left(s,\bar{X}^{t,x;\bar{u},\bar{\xi}}(s),\overline{\varphi}(s),\overline{q}(s),\overline{\Theta}(s),\overline{u}(s)\right),\\
&\text{a.e. } s\in[t,T],\ P\text{-a.s.},
\end{aligned}
\tag{75}
$$

where $\overline{\varphi}(t)$ is defined in Theorem 32. This is easily seen by recalling the fact that V is the viscosity solution of (51):

$$
\overline{p}(s)+\min_{u\in U}\mathcal{G}\left(s,\bar{X}^{t,x;\bar{u},\bar{\xi}}(s),\overline{\varphi}(s),\overline{q}(s),\overline{\Theta}(s),u\right)\geq 0,
$$

which yields (75) under (66).

Remark 35. Clearly, Theorem 32 is expressed in terms of parabolic superjet. One could naturally ask whether a similar result holds for parabolic subjet. The answer was positive for the deterministic case (in terms of the first-order parabolic subjet, see Theorem 3.9 in [77]). Unfortunately, as claimed in Yong and Zhou [77], the answer is that the statement of Theorem 32 is no longer valid whenever the parabolic superjet in (65) is replaced by the parabolic subjet.

Now let us present a non-smooth version of the necessity part of Theorem 32. However, we just have "partial" result.

Theorem 36. *Assume that* (A1)-(A2) *hold. Let* $V \in C([0,T] \times \mathbb{R}^n)$ *be a viscosity solution of the H-J-B equations (51) and let* $(\bar{u}(\cdot), \bar{\xi}(\cdot))$ *be an optimal singular controls. Let* $(\bar{X}^{t,x;\bar{u},\bar{\xi}}(\cdot), \bar{Y}^{t,x;\bar{u},\bar{\xi}}(\cdot), \bar{Z}^{t,x;\bar{u},\bar{\xi}}(\cdot), \bar{u}(\cdot), \bar{\xi}(\cdot))$ *be an admissible pair such that there exist a function* $\varphi \in C^{1,2}([0,T]; \mathbb{R}^n)$ *and a triple*

$$\left(\bar{p}, \bar{q}, \bar{\Theta}\right) \in \left(L^2_{\mathcal{F}_t}([t,T]; \mathbb{R}) \times L^2_{\mathcal{F}_t}([t,T]; \mathbb{R}^n) \times L^2_{\mathcal{F}_t}([t,T]; \mathbb{S}^n)\right)$$

satisfying

$$\begin{cases} \left(\bar{p}(s), \bar{q}(s), \bar{\Theta}(s)\right) \in \mathcal{P}^{2,-} V\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right), \\ \bar{p}(s) G + K \geq 0, \ \text{a.e. } s \in [t,T], \ P\text{-a.s.} \end{cases} \tag{76}$$

Then, it holds that

$$\mathbb{E}\bar{p}(s) \leq -\mathbb{E}\left[\mathcal{G}\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s), \bar{q}(s), \bar{\Theta}(t)\right)\right], \ \text{a.e. } s \in [t,T].$$

Proof. On the one hand, let $s \in [t,T]$ and $\omega \in \Omega$ such that

$$\left(\bar{p}(s), \bar{q}(s), \bar{\Theta}(s)\right) \in \mathcal{P}^{2,-} V\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right).$$

By Lemma 27, we have a test function $\varphi \in C^{1,2}([0,T] \times \mathbb{R}^n)$ with $(s,x) \in [0,T] \times \mathbb{R}^n$ $(p,q,\Theta) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{S}^n$ such that $V - \varphi$ achieves its minimum at $\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right)$ and

$$\left(\frac{\partial \varphi}{\partial t}\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right), D_x\varphi\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right), D^2\varphi\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right)\right) = \left(\bar{p}(s), \bar{q}(s), \bar{\Theta}(s)\right)$$

holds. Then for sufficiently small $\theta > 0$, a.e. $s \in [t,T]$.

$$\begin{aligned}
&\mathbb{E}\left[V\left(s+\theta, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s+\theta)\right) - V\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right)\right] \\
&\geq \mathbb{E}\left[\varphi\left(s+\theta, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s+\theta)\right) - \varphi\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right)\right] \\
&= \mathbb{E}\Bigg[\int_s^{s+\theta} \Big[\frac{\partial \varphi}{\partial t}\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right) + \left\langle D_x\varphi\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right), \bar{b}(r)\right\rangle \\
&\quad + \frac{1}{2}\mathrm{tr}\left\{\bar{\sigma}(r)^* D_{xx}\varphi\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)\bar{\sigma}(r)\right\}\Big]\mathrm{d}r \\
&\quad + \int_{t_0}^{t_0+h} \varphi_x\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right) G\mathrm{d}\bar{\xi}(r) + \sum_{s \leq r \leq s+\theta}\Big\{\varphi\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right) \\
&\quad - \varphi\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right) - \varphi_x\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)\Delta\bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\Big\}\Bigg] \\
&\geq \mathbb{E}\Bigg[\int_s^{s+\theta} \Big[\frac{\partial \varphi}{\partial t}\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right) + \left\langle D_x\varphi\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right), \bar{b}(r)\right\rangle \\
&\quad + \frac{1}{2}\mathrm{tr}\left\{\bar{\sigma}(r)^* D_{xx}\varphi\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right)\bar{\sigma}(r)\right\}\Big]\mathrm{d}r - \int_s^{s+\theta} K\mathrm{d}\bar{\xi}(r)\Bigg].
\end{aligned}$$

The last inequality comes from the derivation in Theorem 32 by means of the condition (76). On the other hand, since $\left(\bar{X}^{t,x;\bar{u},\bar{\xi}}(\cdot), \bar{Y}^{t,x;\bar{u},\bar{\xi}}(\cdot), \bar{Z}^{t,x;\bar{u},\bar{\xi}}(\cdot)\,\bar{u}(\cdot), \bar{\xi}(\cdot)\right)$ is optimal, by DPP of optimality, it yields

$$
\begin{aligned}
&V\left(\tau, \bar{X}^{t,x;\bar{u},\bar{\xi}}(\tau)\right)\\
&= \mathbb{E}^{\mathcal{F}_s^t(\omega)}\left[\Phi\left(\bar{X}^{t,x;\bar{u},\bar{\xi}}(T)\right) + \int_\tau^T \bar{f}(r)\,\mathrm{d}r + \int_\tau^T K\mathrm{d}\bar{\xi}(r)\right], \ \forall \tau \in [t,T], \ P\text{-a.s.},
\end{aligned}
$$

which implies that

$$
\mathbb{E}\left[V\left(s+\theta, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s+\theta)\right) - V\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right)\right] = -\left[\int_s^{s+\theta} \bar{f}(r)\,\mathrm{d}r + \int_\tau^T K\mathrm{d}\bar{\xi}(r)\right]. \tag{77}
$$

Therefore, it follows from (77) that

$$
\begin{aligned}
&\mathbb{E}\left[\bar{p}(s)\right] \le -\mathbb{E}\Big[\tfrac{\partial\varphi}{\partial t}\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right) + \left\langle D_x\varphi\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right), \bar{b}(s)\right\rangle\\
&+\tfrac{1}{2}\mathrm{tr}\left\{\bar{\sigma}(s)^* D_{xx}\varphi\left(s, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right)\bar{\sigma}(s)\right\} - \bar{f}(s)\Big], \ \text{a.e. } s \in [t,T],
\end{aligned}
$$

where

$$
\bar{f}(r) = f\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r), V\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r)\right), \bar{q}(r)\,\sigma\left(r, \bar{X}^{t,x;\bar{u},\bar{\xi}}(r), \bar{u}\right)\right).
$$

We thus complete the proof. □

4.2. Optimal feedback controls

In this subsection, we describe the method to construct optimal feedback controls by the verification Theorem 32. First, let us recall the definition of admissible feedback controls.

Definition 37. A measurable function $(\mathbf{u}, \xi)$ from $[0,T]\times\mathbb{R}^n$ to $U\times[0,\infty)^m$ is called an admissible feedback control pair if for any $(t,x)\in[0,T)\times\mathbb{R}^n$ there is a weak solution $X^{t,x;u,\xi}(\cdot)$ of the following SDE:

$$
\begin{cases}
\mathrm{d}X^{t,x;\mathbf{u},\xi}(r) &= b\left(r, X^{t,x;\mathbf{u},\xi}(r), \mathbf{u}(r)\right)\mathrm{d}r + \sigma\left(r, X^{t,x;\mathbf{u},\xi}(r), \mathbf{u}(r)\right)\mathrm{d}W(r) + G\mathrm{d}\xi_r,\\
\mathrm{d}Y^{t,x;\mathbf{u},\xi}(r) &= -f\left(r, X^{t,x;\mathbf{u},\xi}(r), Y^{t,x;\mathbf{u},\xi}(r), \mathbf{u}(r)\right)\mathrm{d}r + \mathrm{d}M^{t,x;\mathbf{u},\xi}(r) - K\mathrm{d}\xi_r,\\
X^{t,x;\mathbf{u},\xi}(t) &= x, \quad Y^{t,x;\mathbf{u},\xi}(T) = \Phi\left(X^{t,x;\mathbf{u},\xi}(T)\right), \ r\in[t,T],
\end{cases} \tag{78}
$$

where $M^{t,x;\mathbf{u},\xi}$ is an $\mathbb{R}$-valued $\mathbb{F}^{t,x;\mathbf{u},\xi}$-adapted right continuous and left limit martingale vanishing in $t=0$ which is orthogonal to the driving Brownian motion W. Here $\mathbb{F}^{t,x;\mathbf{u},\xi} = \left(\mathcal{F}_s^{X^{t,x;\mathbf{u},\xi}}\right)_{s\in[t,T]}$ is the smallest filtration and generated by $X^{t,x;\mathbf{u},\xi}$, which is such that $X^{t,x;\mathbf{u},\xi}$ is $\mathbb{F}^{t,x;\mathbf{u},\xi}$-adapted. Obviously, $M^{t,x;\mathbf{u},\xi}$ is a part of the solution of BSDE of (78). Simultaneously, we suppose that f satisfies the Lipschitz condition with respect to (x,y,z). An admissible feedback control pair $(\mathbf{u}^\star, \xi^\star)$ is called optimal if

$$
\left(X^\star(\cdot;t,x), Y^\star(\cdot;t,x), \mathbf{u}^\star(\cdot, X^\star(\cdot;t,x)), \xi^\star(\cdot, X^\star(\cdot;t,x))\right)
$$

is optimal for each (t,x) is a solution of (78) corresponding to $(\mathbf{u}^\star, \xi^\star)$.

Theorem 38. *Let* $(\mathbf{u}^{\star}, \xi^{\star})$ *be an admissible feedback control and* $p^{\star}, q^{\star}$, *and* $\Theta^{\star}$ *be measurable functions satisfying* $(p^{\star}(t,x), q^{\star}(t,x), \Theta(t,x)) \in \mathcal{P}^{2,+}V(t,x)$ *for all* $(t,x) \in [0,T] \times \mathbb{R}^n$. *If*

$$\begin{aligned} & p^{\star}(t,x) + \mathcal{G}(t,x,V(t,x), q^{\star}(t,x), \Theta^{\star}(t,x), \mathbf{u}^{\star}(t,x)) \\ = & \inf_{(p,q,\Theta,u) \in \mathcal{P}^{2,+}v(t,x) \times U} [p + \mathcal{G}(t,x,V(t,x), q, \Theta, u)] = 0 \end{aligned} \tag{79}$$

and $q^{\star}(t,x) G + K \geq 0$ *for all* $(t,x) \in [0,T] \times \mathbb{R}^n$, *then* $(\mathbf{u}^{\star}, \xi^{\star})$ *is singular optimal control pair.*

Proof. From Theorem 32, we get the desired result. □

Remark 39. In FBSDEs (78), $Y^{t,x;u,\xi}(\cdot)$ is actually determined by $\left(X^{t,x;u,\xi}(\cdot), u(\cdot), \xi(\cdot)\right)$. Hence, we need to investigate the conditions imposed in Theorem 32 to ensure the existence and uniqueness of $X^{t,x;u,\xi}(\cdot)$ in law and the measurability of the multifunctions $(t,x) \to \mathcal{P}^{2,+}V(t,x)$ to obtain $(p^{\star}(t,x), q^{\star}(t,x), \Theta(t,x)) \in \mathcal{P}^{2,+}V(t,x)$ that minimizes (79). This can be done by virtue of the celebrated Filippov's Lemma (cf. [77]).

4.3. The connection between DPP and MP

In Section 3, we have obtained the first and second order adjoint equations. In this part, we shall investigate the connection between the general DPP and the MP for such singular controls problem without the assumption that the value is sufficient smooth. By associated adjoint equations and delicate estimates, it is possible to establish the set inclusions among the super- and sub-jets of the value function and the first-order and second-order adjoint processes as well as the generalized Hamiltonian function.

Theorem 40. *Assume that* (A1)-(A2) *are in force. Suppose that* $(\bar{u}, \bar{\xi})$ *be a singular optimal controls,* $V(\cdot,\cdot)$ *is a value function, and* $\left(\bar{X}^{t,x;\bar{u},\bar{\xi}}(\cdot), \bar{Y}^{t,x;\bar{u},\bar{\xi}}(\cdot), \bar{Z}^{t,x;\bar{u},\bar{\xi}}(\cdot), \bar{u}(\cdot), \bar{\xi}(\cdot)\right)$ *is optimal trajectory. Let* $(p,q) \in \mathcal{S}^2(0,T;\mathbb{R}^n) \times \mathcal{M}^2(0,T;\mathbb{R}^n)$ *and* $(P,Q) \in \mathcal{S}^2(0,T;\mathbb{R}^{n\times n}) \times \mathcal{M}^2(0,T;\mathbb{R}^{n\times n})$ *be the adjoint equations (25), (26), respectively. Then, we have*

$$\begin{aligned} P\left\{K_{(i)} + p(t) G_{(i)}(t) \geq 0,\ t \in [0,T],\ \forall i\right\} = 1, \\ \{p(s)\} \times [P(s), \infty) \subseteq \mathcal{P}^{2,+}V\left(t, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right), \\ \mathcal{P}^{2,-}V\left(t, \bar{X}^{t,x;\bar{u},\bar{\xi}}(s)\right) \subseteq \{p(s)\} \times [-\infty, P(s)),\ \textit{a.e. } s \in [t,T],\ P\textit{-a.s.} \end{aligned} \tag{80}$$

Proof. From Theorem 3 and Proposition 8, we get the first part of (80). From Theorem 3.1 in Nie, Shi and Wu [58], we get the second and third results of (80). □

5. Concluding remarks

In this paper, on the one hand, we have derived a second order pointwise necessary condition for singular optimal control in classical sense of FBSDEs with convex control domain by means of the variation equations and two adjoint equations, which is separately extends the work by Zhang and Zhang [80] to stochastic recursive case, and Hu [40] to pointwise case in the framework of Malliavin calculus. A new necessary condition for singular control has been obtained. Moreover, we investigate the verification theorem for optimal controls via viscosity solution and establish the connection between the adjoint equations and value function also in viscosity solution sense.

There are still several interesting topics should be scheduled as follows:

- As an important issue, the *existence* of optimal singular controls has never been exploited. Haussmann and Suo [38] apply the compactification method to study the classical and singular control problem of

Itô's type of stochastic differential equation, where the problem is reformulated as a martingale problem on an appropriate canonical space after the relaxed form of the classical control is introduced. Under some mild continuity assumptions on the data, they obtain the existence of optimal control by purely probabilistic arguments. Note that, in the framework of BSDE with singular control, the trajectory of Y seems to be a càdlàg process (from French, for right continuous with left hand limits). Hence, we may consider Y in some space with appropriate topologies, for instance, Skorokhod M_1 topology or Meyer-Zheng topology (see [34]) to obtain the convergence of probability measures deduced by Y involving relaxed control. Related work from the technique of PDEs can be seen in [4,19] references therein. From Wang [72], one may construct the optimal control via the existence of diffusion with refections (see [23]). However, it is interesting to extend this result to FBSDEs.

- The matrices K, G are deterministic. It is also interesting to extend this restriction to time varying matrices, even the generator b, σ, f involving the singular control. Whenever the coefficients are random, the H-J-B inequality will become stochastic PDEs. No doubt, stochastic viscosity solution will be applied. For this direction, reader can refer to Buckdahn, Ma [17,18], Peng [65] and Qiu [68].
- As for the general cases, *i.e.*, the control regions are assumed to be non-convex and both the drift and diffusion terms depend on the control variable. Indeed, such a mathematical model, from view point of application, is more reasonable and urgent in many real-life problems (for instance, some finance models in which the controls may impact the uncertainty, *etc.*). In near future, we shall remove the condition of convex control region, employing the idea developed by Zhang et al. [81]. It is worth mentioning that the analysis in [81] is much more complicated. Some new and useful tools, such as the multilinear function valued stochastic processes, the BSDE for these processes are introduced. Hence, it will be interesting to borrow these tools to investigate the singular optimal controls problems for FBSDEs, which will definitely promote and enrich the theories of FBSDEs.

Acknowledgments

The author is very grateful to the referee and the AE, Prof. Marc Quincampoix, for their fruitful comments and suggestions, which make our paper more readable.

Appendix A. Proofs of lemmas

Proof of Lemma 4. We first prove the continuity of solution depending on parameter.

Set

$$\begin{aligned}\hat{X}^{\alpha}(s) &= X^{0,x;\bar{u},\xi^{\alpha}}(s) - X^{0,x;\bar{u},\bar{\xi}}(s),\\ \hat{Y}^{\alpha}(s) &= Y^{0,x;\bar{u},\xi^{\alpha}}(s) - Y^{0,x;\bar{u},\bar{\xi}}(s),\\ \hat{Z}^{\alpha}(s) &= Z^{0,x;\bar{u},\xi^{\alpha}}(s) - Z^{0,x;\bar{u},\bar{\xi}}(s),\ \ s \in [0,T].\end{aligned}$$

It can be shown that

$$\left(\hat{X}^{\alpha}(t), \hat{Y}^{\alpha}(t), \hat{Z}^{\alpha}(t)\right) \text{ converges to 0 in } \mathcal{N}^2[0,T] \text{ as } \alpha \to 0 \tag{81}$$

by standard estimates and the Burkholder-Davis-Gundy inequality, so we omit it.

Next, set

$$\Delta X^{\alpha}(s) = \frac{\hat{X}^{\alpha}(s)}{\alpha} - x^1(s),\ \Delta Y^{\alpha}(s) = \frac{\hat{Y}^{\alpha}(s)}{\alpha} - y^1(s),\ \Delta Z^{\alpha}(s) = \frac{\hat{Z}^{\alpha}(s)}{\alpha} - z^1(s).$$

Note that (19) has be obtained in [3]. We will prove (20) and (21).

Then,

$$\begin{cases} \mathrm{d}\Delta X^{\alpha}(s) &= \Delta b^{\alpha}(s)\,\mathrm{d}s + \Delta\sigma^{\alpha}(s)\,\mathrm{d}W(s), \\ -\mathrm{d}\Delta Y^{\alpha}(s) &= \Delta f^{\alpha}(s)\,\mathrm{d}s - \Delta Z^{\alpha}(s)\,\mathrm{d}W(s), \\ \Delta X^{\alpha}(0) &= 0,\ \Delta Y^{\alpha}(T) = \Delta\Phi^{\alpha}(T), \qquad 0 \le t \le s \le T, \end{cases} \tag{82}$$

where

$$\begin{aligned}
\Delta b^{\alpha}(s) &= \frac{b\left(s, X^{0,x;\bar{u},\xi^{\alpha}}(s), \bar{u}(s)\right) - b\left(s, X^{0,x;\bar{u},\bar{\xi}}(s), \bar{u}(s)\right)}{\alpha} - \bar{b}_x(t)\,x^1(t), \\
\Delta\sigma^{\alpha}(s) &= \frac{\sigma\left(s, X^{0,x;\bar{u},\xi^{\alpha}}(s), \bar{u}(s)\right) - \sigma\left(s, X^{0,x;\bar{u},\bar{\xi}}(s), \bar{u}(s)\right)}{\alpha} - \bar{\sigma}_x(t)\,x^1(t), \\
\Delta f^{\alpha}(s) &= \frac{1}{\alpha}\Bigg[f\left(s, X^{0,x;\bar{u},\xi^{\alpha}}(s), Y^{0,x;\bar{u},\xi^{\alpha}}(s), Z^{0,x;\bar{u},\xi^{\alpha}}(s), \bar{u}(s)\right) \\
&\quad - f\left(s, X^{0,x;\bar{u},\bar{\xi}}(s), Y^{0,x;\bar{u},\bar{\xi}}(s), Z^{0,x;\bar{u},\bar{\xi}}(s), \bar{u}(s)\right)\Bigg] \\
&\quad - \bar{f}_x(s)\,x^1(s) - \bar{f}_y(s)\,y^1(s) - \bar{f}_z(s)\,z^1(s), \\
\Delta\Phi^{\alpha}(T) &= \frac{\Phi\left(X^{0,x;\bar{u},\xi^{\alpha}}(T)\right) - \Phi\left(X^{0,x;\bar{u},\bar{\xi}}(T)\right)}{\alpha} - \Phi_x\left(X^{0,x;\bar{u},\bar{\xi}}(T)\right)x^1(T).
\end{aligned}$$

Simple calculation yields

$$\begin{aligned}
&\Delta f^{\alpha}(s) \\
&= \int_0^1 f_x\left(s, \lambda X^{0,x;\bar{u},\xi^{\alpha}}(s) + (1-\lambda) X^{0,x;\bar{u},\bar{\xi}}(s), Y^{0,x;\bar{u},\xi^{\alpha}}(s), Z^{0,x;\bar{u},\xi^{\alpha}}(s), \bar{u}(s)\right)\Delta X^{\alpha}(s)\,\mathrm{d}\lambda \\
&+ \int_0^1 f_y\left(s, X^{0,x;\bar{u},\bar{\xi}}(s), \lambda Y^{0,x;\bar{u},\xi^{\alpha}}(s) + (1-\lambda) Y^{0,x;\bar{u},\bar{\xi}}(s), Z^{0,x;\bar{u},\xi^{\alpha}}(s), \bar{u}(s)\right)\Delta Y^{\alpha}(s)\,\mathrm{d}\lambda \\
&+ \int_0^1 f_z\left(s, X^{0,x;\bar{u},\bar{\xi}}(s), Y^{0,x;\bar{u},\bar{\xi}}(s), \lambda Z^{0,x;\bar{u},\xi^{\alpha}}(s) + (1-\lambda) Z^{0,x;\bar{u},\bar{\xi}}(s), \bar{u}(s)\right)\Delta Z^{\alpha}(s)\,\mathrm{d}\lambda \\
&+ \Delta\rho^{\alpha}(s),
\end{aligned}$$

where

$$\begin{aligned}
&\Delta\rho^{\alpha}(s) \\
&= \Bigg[\int_0^1 f_x\left(s, \lambda X^{0,x;\bar{u},\xi^{\alpha}}(s) + (1-\lambda) X^{0,x;\bar{u},\bar{\xi}}(s), Y^{0,x;\bar{u},\xi^{\alpha}}(s), Z^{0,x;\bar{u},\xi^{\alpha}}(s), \bar{u}(s)\right)\mathrm{d}\lambda \\
&- \bar{f}_x(s)\,x^1(s)\,\mathrm{d}s\Bigg] \\
&+ \left[\int_0^1 f_y\left(s, X_s^{0,x;\bar{u},\bar{\xi}}, \lambda Y_s^{0,x;\bar{u},\xi^{\alpha}} + (1-\lambda) Y^{0,x;\bar{u},\bar{\xi}}(s), Z_s^{0,x;\bar{u},\xi^{\alpha}}, \bar{u}(s)\right)\mathrm{d}\lambda - \bar{f}_y(s)\right] y^1(s)\,\mathrm{d}s
\end{aligned}$$

$$+\left[\int_0^1 f_z\left(s, X_s^{0,x;\bar{u},\bar{\xi}}, Y_s^{0,x;\bar{u},\bar{\xi}}, \lambda Z_s^{0,x;\bar{u},\xi^\alpha} + (1-\lambda) Z_s^{0,x;\bar{u},\bar{\xi}}, \bar{u}(s)\right) \mathrm{d}\lambda - \bar{f}_z(s)\right] z^1(s)\,\mathrm{d}s$$

and

$$\begin{aligned}
&\Delta\Phi^\alpha(T)\\
&=\int_0^1 \Phi_x\left(\lambda X^{0,x;\bar{u},\xi^\alpha}(T) + (1-\lambda) X^{0,x;\bar{u},\bar{\xi}}(T)\right) \Delta X^\alpha(s)\,\mathrm{d}\lambda\\
&\quad+\left[\int_0^1 \Phi_x\left(\lambda X^{0,x;\bar{u},\xi^\alpha}(T) + (1-\lambda) X^{0,x;\bar{u},\bar{\xi}}(T)\right)\mathrm{d}\lambda - \Phi_x\left(X_T^{0,x;\bar{u},\bar{\xi}}\right)\right] x^1(T)\,\mathrm{d}s.
\end{aligned}$$

From classical theory of BSDE, one can show

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|x^1(t)\right|^2\right] < \infty, \mathbb{E}\left[\sup_{0\le t\le T}\left|y^1(t)\right|^2 \mathrm{d}t\right] < \infty, \mathbb{E}\left[\int_0^T \left|z^1(t)\right|^2 \mathrm{d}t\right] < \infty. \tag{83}$$

By using (81) and (83), the dominated convergence theorem, Lemma 1 and Gronwall's lemma, we get the desired result by letting $\alpha \to 0$. □

Proof of Lemma 33. From (62) and (6) in Gozzi et al. in [37], we have that if $(p,q,P) \in \mathcal{P}^{2,+}V(t,x)$, then

$$\begin{aligned}
&V\left(t+h, X^{t,x;\bar{u},\bar{\xi}}(t+h)\right) - V\left(t, X^{t,x;\bar{u},\bar{\xi}}(t)\right) \le C\left(1+\left|X^{t,x;\bar{u},\bar{\xi}}(t)\right|^m\right)h\\
&+\left\langle q(t), X^{t,x;\bar{u},\bar{\xi}}(t+h) - X^{t,x;\bar{u},\bar{\xi}}(t)\right\rangle + C_0\left|X^{t,x;\bar{u},\bar{\xi}}(t+h) - X^{t,x;\bar{u},\bar{\xi}}(t)\right|^2\\
&= I_1 + I_2 + I_3.
\end{aligned}$$

We shall deal with I_1, I_2, I_3, separately. For I_1, we have $\mathbb{E}\left(1+\left|X^{t,x;\bar{u},\bar{\xi}}(t+h)\right|^m\right)h \le C(1+|x|^m)h$, by classical estimate and the assumption $\mathbb{E}\left[|\xi_T|^2\right] < \infty$. For I_2, from (7) in [37] and Hölder inequality, we have

$$\begin{aligned}
&\mathbb{E}\left\langle q(t), X^{t,x;\bar{u},\bar{\xi}}(t+h) - X^{t,x;\bar{u},\bar{\xi}}(t)\right\rangle\\
&\le C\left(\mathbb{E}\left[\left(1+\left|X^{t,x;\bar{u},\bar{\xi}}(t)\right|^{m_1}\right)^2\right]\right)^{\frac{1}{2}}\left\{\left(\mathbb{E}\left[\left|\int_t^{t+h} b\left(r, X^{0,x;\bar{u},\xi^\alpha}(r), \bar{u}(r)\right)\right|^2\right]\right)^{\frac{1}{2}}\right.\\
&\quad\left.+\left(\mathbb{E}\left[\left|\int_t^{t+h} G(r)\,\mathrm{d}\bar{\xi}(r)\right|^2\right]\right)^{\frac{1}{2}}\right\}\\
&\le C\left(\mathbb{E}\left[(1+|x|^{m_1})^2\right]\right)^{\frac{1}{2}},
\end{aligned}$$

since $\mathbb{E}\left[|\xi_T|^2\right] < \infty$ and the fact $\left(1+|x|^2\right)^{\frac{1}{2}} \le 1+|x|$.

Finally,

$$C_0\mathbb{E}\left|X^{t,x;\bar{u},\bar{\xi}}(t+h)-X^{t,x;\bar{u},\bar{\xi}}(t)\right|^2 \leq C_0\mathbb{E}\left[\left|\int_t^{t+h} b\left(r,X^{0,x;\bar{u},\xi^\alpha}(r),\bar{u}(r)\right)\right|^2\right]$$

$$+C_0\mathbb{E}\left[\left|\int_t^{t+h}\sigma\left(r,X^{0,x;\bar{u},\xi^\alpha}(r),\bar{u}(r)\right)\right|^2\right]+C_0\mathbb{E}\left[\left|\int_t^{t+h}G(r)\,\mathrm{d}\bar{\xi}(r)\right|^2\right]$$

$$\leq C\left(1+|x|^2\right)\left(2h^2+h\right).$$

By Itô isometry and classical estimate on SDE, we complete the proof. □

References

[1] L. Alvarez, A class of solvable singular stochastic control problems, Stoch. Stoch. Rep. 67 (1999) 83–122.
[2] L. Alvarez, Singular stochastic control, linear diffusions, and optimal stopping: a class of solvable problems, SIAM J. Control Optim. 39 (2001) 1697–1710.
[3] S. Bahlali, B. Mezerdi, B. Djehiche, The relaxed stochastic maximum principle in singular control of diffusions, SIAM J. Control Optim. 46 (2007) 427–444.
[4] K. Bahlali, B. Gherbal, B. Mezerdi, Existence of optimal controls for systems driven by FBSDEs, Syst. Control Lett. 60 (2011) 344–349.
[5] F. Baldursson, Singular stochastic control and optimal stopping, Stochastics 21 (1987) 1–40.
[6] F. Baldursson, I. Karatzas, Irreversible investment and industry equilibrium, Finance Stoch. 1 (1997) 69–89.
[7] J. Bather, H. Chernoff, Sequential decisions in the control of a spaceship, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. III: Physical Sciences, University of California Press, Berkeley, CA, 1967, pp. 181–207.
[8] D.J. Bell, D.H. Jacobson, Singular Optimal Control Problems, Math. Sci. Eng., vol. 117, Academic Press, New York, 1975.
[9] V.E. Beněs, L.A. Shepp, H.S. Witzsenhausen, Some solvable stochastic control problems, Stochastics 4 (1980) 39–83.
[10] J.M. Bismut, Théorie Probabiliste du Contrôle des Diffusions, Memoirs of the American Mathematical Society, vol. 176, American Mathematical Society, Providence, RI, 1973.
[11] J. Bismut, An introductory approach to duality in optimal stochastic control, SIAM Rev. 20 (1978) 62–78.
[12] F. Boetius, Bounded variation singular stochastic control and associated Dynkin game, in: M. Kohlmann, S. Tang (Eds.), Mathematical Finance, in: Trends Math., Birkhäuser, Basel, 2001, pp. 111–120.
[13] F. Boetius, Singular stochastic control and its relations to Dynkin game and rntry-exit problems, Ph.D. thesis, University of Konstanz, 2001.
[14] F. Boetius, M. Kohlmann, Connections between optimal stopping and singular stochastic control, Stoch. Process. Appl. 77 (1998) 253–281.
[15] J.F. Bonnans, F.J. Silva, First and second order necessary conditions for stochastic optimal control problems, Appl. Math. Optim. 65 (2012) 403–439.
[16] Ph. Briand, B. Delyon, Y. Hu, E. Pardoux, L. Stoica, L^p solutions of backward stochastic differential equations, Stoch. Process. Appl. 108 (2003) 109–129.
[17] R. Buchdahn, J. Ma, Stochastic viscosity solutions for nonlinear stochastic. Part I, Stoch. Process. Appl. 93 (2001) 181–204.
[18] R. Buckdahn, J. Ma, Pathwise stochastic Taylor expansions and stochastic viscosity solutions for fully nonlinear stochastic PDEs, Ann. Probab. 30 (2002) 1131–1171.
[19] R. Buckdahn, B. Labed, C. Rainer, L. Tamer, Existence of an optimal control for stochastic control systems with nonlinear cost functional, Int. J. Probab. Stoch. Process. 82 (3) (2010).
[20] A. Cadenillas, U. Haussmann, The stochastic maximum principle for a singular control problem, Stoch. Stoch. Rep. 49 (1994) 211–237.
[21] Z. Chen, L. Epstein, Ambiguity, risk, and asset returns in continuous time, Econometrica 70 (2002) 1403–1443.
[22] M.B. Chiarolla, U.G. Haussmann, Optimal control of inflation: central bank problem, SIAM J. Control Optim. 36 (1998) 1099–1132.
[23] P. Chow, J.-L. Menaldi, M. Robin, Additive control of stochastic linear systems with finite horizon, SIAM J. Control Optim. 23 (1985) 858–899.
[24] M.G. Crandall, H. Ishii, P.-L. Lions, User's guide to viscosity solutions of second order partial differential equations, Bull. Am. Math. Soc. (N.S.) 27 (1992) 1–67.
[25] K.R. Dahl, B. Øksendal, Singular recursive utility, Stochastics 89 (6–7) (2017) 994–1014.
[26] M.H.A. Davis, A.R. Norman, Portfolio selection with transaction costs, Math. Oper. Res. 15 (4) (1990) 676–713.
[27] F. Dufour, B. Miller, Maximum principle for singular stochastic control problems, SIAM J. Control Optim. 45 (2) (2006) 668–698.
[28] N. El Karoui, I. Karatzas, Probabilistic aspects of finite-fuel, reflected follower problems, Acta Appl. Math. 11 (1988) 223–258.

[29] N. El Karoui, I. Karatzas, A new approach to the Skorohod problem, and its applications, Stoch. Stoch. Rep. 34 (1991) 57–82.
[30] N. El Karoui, S. Peng, M.C. Quenez, Backward stochastic differential equations in finance, Math. Finance 7 (1997) 1–71.
[31] W.H. Fleming, R.W. Rishel, Deterministic and Stochastic Optimal Control, Springer-Verlag, New York, 1975.
[32] W. Fleming, H. Soner, Controlled Markov Processes and Viscosity Solutions, Springer-Verlag, New York, 1993.
[33] H. Frankowska, D. Tonon, Pointwise second-order necessary optimality conditions for the Mayer problem with control constraints, SIAM J. Control Optim. 51 (2013) 3814–3843.
[34] G. Fu, U. Horst, Mean field games with singular controls, SIAM J. Control Optim. 55 (6) (2017) 3833–3868.
[35] R. Gabasov, F.M. Kirillova, High order necessary conditions for optimality, SIAM J. Control Optim. 10 (1972) 127–168.
[36] B.S. Goh, Necessary conditions for singular extremals involving multiple control variables, SIAM J. Control Optim. 4 (1966) 716–731.
[37] F. Gozzi, A. Święch, X. Zhou, Erratum: "A corrected proof of the stochastic verification theorem within the framework of viscosity solutions", SIAM J. Control Optim. 48 (6) (2010) 4177–4179.
[38] U. Haussmann, W. Suo, Singular optimal stochastic controls I: existence, SIAM J. Control Optim. 33 (3) (1995) 916–936.
[39] U. Haussmann, W. Suo, Singular optimal stochastic controls II: dynamic programming, SIAM J. Control Optim. 33 (3) (1995) 937–959.
[40] M. Hu, Stochastic global maximum principle for optimization with recursive utilities, Probab. Uncertain. Quant. Risk 2 (2017) 1.
[41] M. Hu, S. Ji, X. Xue, A global stochastic maximum principle for fully coupled forward-backward stochastic systems, SIAM J. Control Optim. 56 (6) (2018) 4309–4335.
[42] S. Ji, X.Y. Zhou, A maximum principle for stochastic optimal control with terminal state constrains, and its applications, Commun. Inf. Syst. 6 (2006) 321–337.
[43] I. Karatzas, The monotone follower problem in stochastic decision theory, Appl. Math. Optim. 7 (1981) 175–189.
[44] I. Karatzas, A class of singular control problems, Adv. Appl. Probab. 15 (1983) 225–254.
[45] I. Karatzas, S. Shreve, Connections between optimal stopping and singular stochastic control I. Monotone follower problems, SIAM J. Control Optim. 22 (1984) 856–877.
[46] I. Karatzas, S. Shreve, Connections between optimal stopping and singular stochastic control II. Reflected follower problems, SIAM J. Control Optim. 23 (1985) 433–451.
[47] I. Karatzas, S. Shreve, Equivalent models for finite-fuel stochastic control, Stochastics 18 (1986) 245–276.
[48] I. Karatzas, H. Wang, A barrier option of American type, Appl. Math. Optim. 42 (2000) 259–279.
[49] H.-W. Knobloch, Higher Order Necessary Conditions in Optimal Control Theory, Lecture Notes in Comput. Sci., vol. 34, Springer-Verlag, New York, 1981.
[50] T. Kobila, A class of solvable stochastic investment problems involving singular controls, Stoch. Stoch. Rep. 43 (1993) 29–63.
[51] M. Kohlmann, X.Y. Zhou, Relationship between backward stochastic differential equations and stochastic controls: a linear-quadratic approach, SIAM J. Control Optim. 38 (2000) 1392–1407.
[52] A.J. Krener, The high order maximal principle and its application to singular extremals, SIAM J. Control Optim. 15 (1977) 256–293.
[53] N. Krylov, Controlled Diffusion Processes, Springer-Verlag, New York, 1980.
[54] H.J. Kushner, Necessary conditions for continuous parameter stochastic optimization problems, SIAM J. Control Optim. 10 (1972) 550–565.
[55] A. Lim, X.Y. Zhou, Linear-quadratic control of backward stochastic differential equations, SIAM J. Control Optim. 40 (2001) 450–474.
[56] L.F. Martins, H.J. Kushner, Routing and singular control for queueing networks in heavy traffic, SIAM J. Control Optim. 28 (1990) 1209–1233.
[57] J.-L. Menaldi, M. Taksar, Optimal correction problem of a multidimensional stochastic system, Automatica 25 (1989) 223–232.
[58] T. Nie, J. Shi, Z. Wu, Connection between MP and DPP for stochastic recursive optimal control problems: viscosity solution framework in the general case, SIAM J. Control Optim. 55 (5) (2017) 3258–3294.
[59] D. Nualart, The Malliavin Calculus and Related Topics, Probability and Its Applications, Springer-Verlag, New York, Berlin, 1995.
[60] D. Nualart, E. Pardoux, Stochastic calculus with anticipating integrands, Probab. Theory Relat. Fields 78 (1988) 535–581.
[61] D. Ocone, Malliavin's calculus and stochastic integral representation of functionals of diffusion processes, Stochastics 12 (1988) 161–185.
[62] H. Pagès, D. Possamaï, A mathematical treatment of bank monitoring incentives, Finance Stoch. 18 (1) (2014) 39–73.
[63] E. Pardoux, S. Peng, Adapted solution of a backward stochastic differential equation, Syst. Control Lett. 14 (1990) 55–61.
[64] S. Peng, A general stochastic maximum principle for optimal control problems, SIAM J. Control Optim. 28 (1990) 966–979.
[65] S. Peng, Stochastic Hamilton-Jacobi-Bellman equations, SIAM J. Control Optim. 30 (1992) 284–304.
[66] S. Peng, Backward stochastic differential equations and applications to optimal control, Appl. Math. Optim. 27 (1993) 125–144.
[67] S. Peng, BSDE and stochastic optimizations, in: J. Yan, S. Peng, S. Fang, L. Wu (Eds.), Topics in Stochastic Analysis, Science Press, Beijing, 1997, Chapter 2 (in Chinese).
[68] J. Qiu, Viscosity solutions of stochastic Hamilton–Jacobi–Bellman equations, SIAM J. Control Optim. 56 (5) (2018) 3708–3730.
[69] J. Shi, Z. Wu, The maximum principle for fully coupled forward-backward stochastic control system, Acta Autom. Sin. 32 (2006) 161–169.
[70] S.E. Shreve, J.P. Lehoczky, D.P. Gaver, Optimal consumption for general diffusions with absorbing and reflecting barriers, SIAM J. Control Optim. 22 (1984) 55–75.

[71] S. Tang, A second-order maximum principle for singular optimal stochastic controls, Discrete Contin. Dyn. Syst., Ser. B 14 (2010) 1581–1599.
[72] B. Wang, Singular control of stochastic linear systems with recursive utility, Syst. Control Lett. 51 (2004) 105–122.
[73] Z. Wu, Maximum principle for optimal control problem of fully coupled forward-backward stochastic systems, Syst. Sci. Math. Sci. 11 (1998) 249–259.
[74] Z. Wu, A general maximum principle for optimal control of forward-backward stochastic systems, Automatica 49 (2013) 1473–1480.
[75] W. Xu, Stochastic maximum principle for optimal control problem of forward and backward system, J. Aust. Math. Soc. Ser. B 37 (1995) 172–185.
[76] J. Yong, Optimality variational principle for controlled forward-backward stochastic differential equations with mixed initial-terminal conditions, SIAM J. Control Optim. 48 (2010) 4119–4156.
[77] J. Yong, X.Y. Zhou, Stochastic Controls. Hamiltonian Systems and HJB Equations, Springer-Verlag, New York, 1999.
[78] L. Zhang, Stochastic verification theorem of forward–backward controlled systems for viscosity solutions, Syst. Control Lett. 61 (2012) 649–654.
[79] L. Zhang, Singular optimal controls of stochastic recursive systems and Hamilton–Jacobi–Bellman inequality, J. Differ. Equ. 266 (10) (2019) 6383–6425.
[80] H. Zhang, X. Zhang, Pointwise second-order necessary conditions for stochastic optimal controls, Part I: the case of convex control constraint, SIAM J. Control Optim. 53 (4) (2015) 2267–2296.
[81] H. Zhang, X. Zhang, Pointwise second-order necessary conditions for stochastic optimal controls, Part II: the general case, SIAM J. Control Optim. 55 (5) (2017) 2841–2875.