1-1-2010

Optimality Variational Principle For Controlled Forward-Backward Stochastic Differential Equations With Mixed Initial-Terminal Conditions

Jiongmin Yong
University of Central Florida

Find similar works at: https://stars.library.ucf.edu/facultybib2010
University of Central Florida Libraries http://library.ucf.edu

This Article is brought to you for free and open access by the Faculty Bibliography at STARS. It has been accepted for inclusion in Faculty Bibliography 2010s by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

Recommended Citation
Yong, Jiongmin, "Optimality Variational Principle For Controlled Forward-Backward Stochastic Differential Equations With Mixed Initial-Terminal Conditions" (2010). Faculty Bibliography 2010s. 990.
https://stars.library.ucf.edu/facultybib2010/990
OPTIMALITY VARIATIONAL PRINCIPLE FOR CONTROLLED
FORWARD-BACKWARD STOCHASTIC DIFFERENTIAL
EQUATIONS WITH MIXED INITIAL-TERMINAL CONDITIONS

JIONGMIN YONG†

In memory of Professor Leonard D. Berkovitz

Abstract. An optimal control problem for general coupled forward-backward stochastic differential equations (FBSDEs) with mixed initial-terminal conditions is considered. The control domain is not assumed to be convex, and the control appears in the diffusion coefficient of the forward equation. Necessary conditions of Pontraygin’s type for the optimal controls are derived by means of spike variation techniques.

Key words. forward-backward stochastic differential equations, mixed initial-terminal conditions, optimal stochastic control, optimality variational principle, spike variation technique

AMS subject classifications. 93E20, 49K45, 60H10

DOI. 10.1137/090763287

1. Introduction. Throughout this paper, we let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete filtrated probability space on which a standard one-dimensional Brownian motion $W(\cdot)$ is defined, with $\mathcal{F} = \{\mathcal{F}_t\}_{t \geq 0}$ being its natural filtration augmented by all the $\mathbb{P}$-null sets. Consider a controlled stochastic differential equation (SDE) of Itô type,

\begin{align*}
\begin{aligned}
&dX(t) = B(t, X(t), v(t))dt + \Sigma(t, X(t), v(t))dW(t), \quad t \in [0, T], \\
&X(0) = X_0,
\end{aligned}
\end{align*}

(1.1)

with cost functional

\begin{align*}
J_0(X_0, v(\cdot)) = \mathbb{E} \left[ \int_0^T G(t, X(t), v(t))dt + F(X_0, X(T)) \right],
\end{align*}

(1.2)

where $v(\cdot)$ is a control process taking values in some metric space $V$, and $X(\cdot)$, the strong solution of (1.1), is the corresponding state process taking values in $\mathbb{R}^\ell$, with initial state $X_0 \in K \subseteq \mathbb{R}^\ell$ (not necessarily fixed, $K$ is some subset of $\mathbb{R}^\ell$, which is allowed to be $\mathbb{R}^\ell$). In the above, $B, \Sigma, G, F$ are given maps and $\mathbb{E}$ stands for the expectation with respect to the given (objective) probability measure $\mathbb{P}$. Let $\mathcal{V}[0, T]$ be the set of all $\mathbb{F}$-progressively measurable processes valued in $V$, defined on $[0, T]$. Under some mild conditions, for any $(X_0, v(\cdot)) \in \mathbb{R}^\ell \times \mathcal{V}[0, T]$, (1.1) admits a unique strong solution $X(\cdot) \equiv X(\cdot; X_0, v(\cdot))$. Then we can pose the following optimal control problem.

**Problem (A).** Find a pair $(\bar{X}_0, \bar{v}(\cdot)) \in K \times \mathcal{V}[0, T]$ such that

\begin{align*}
J_0(\bar{X}_0, \bar{v}(\cdot)) = \inf_{(X_0, v(\cdot)) \in K \times \mathcal{V}[0, T]} J_0(X_0, v(\cdot)).
\end{align*}

(1.3)

*Received by the editors July 6, 2009; accepted for publication (in revised form) March 1, 2010; published electronically May 14, 2010. This work is supported in part by NSF grant DMS-0604309.\ *

http://www.siam.org/journals/sicon/48-6/76328.html

†Department of Mathematics, University of Central Florida, Orlando, FL 32816 (jyong@mail.ucf.edu).
If \((\tilde{X}_0, \tilde{v}(\cdot)) \in K \times \mathcal{V}[0, T]\) satisfies the above, we call \(\tilde{v}(\cdot)\) an optimal control process of Problem (A) and the corresponding state process \(\tilde{X}(\cdot)\) is called the corresponding optimal state process.

Note that in a typical classical optimal stochastic control problem (see [14], [40], [22]), one usually has a fixed initial state \(X(0) = X_0 \in \mathbb{R}^n\); therefore, the term \(F(X_0, X(T))\) in (1.2) is replaced by \(F(X(T))\), whereas, if \(X_0 \in K\) is not fixed, the term \(F(X_0, X(T))\) could represent some kind of penalty/reward on the pair \((X_0, X(T))\) jointly, which is standard in optimal control theory of ordinary differential equations (ODEs) (see [3], for example). We therefore naturally include such a term in the stochastic case here, allowing \(X_0\) to be not necessarily fixed. Functional (1.2) essentially represents the preferences of the controller: Between any two pairs of initial state and control, the controller prefers the one that has a smaller value of the cost functional \(J_0(\cdot)\). Hence, functions \(G\) and \(F\) can be regarded as representations of the preferences for the controller. Consequently, they could be very subjective due to the fact that different groups of people might have quite different preferences. One can regard \(-J_0(X_0, v(\cdot))\) as a sort of (expected) utility resulting from taking the pair \((X_0, v(\cdot))\), which is some kind of generalization of the von-Neumann–Morgenstern expected utility [23].

Since the early 1950s, there have appeared the well-known Allais [1] and Ellsberg [13] paradoxes, which indicated that the von-Neumann–Morgenstern linear expected utility theory (by which we mean that in (1.2) the (linear) expectation \(\mathbb{E}\) is used) does not necessarily well represent people’s preferences. Naturally, one has been trying to replace \(\mathbb{E}\) by some kind of nonlinear expectation. An interesting one is the so-called Choquet integral (with respect to some normalized capacity or fuzzy measure which is supposed to be subjective). This leads to the Choquet expected utility theory (see [8], and the references cited therein). One can use this theory to explain the Allais and Ellsberg paradoxes. In our context, it amounts to replacing \(\mathbb{E}\) in (1.2) by a Choquet integral. It seems possible to formulate stochastic optimal control problems with a Choquet expected utility as a payoff functional. However, to our best knowledge, the Choquet integral seems to lack a good “dynamic” feature, which prevents many stochastic analysis tools from being applicable.

On the other hand, Peng [26] introduced a notion of nonlinear expectation (called \(g\)-expectation), denoted by \(\mathbb{E}_g\), based on the theory of backward stochastic differential equations (BSDEs). Such a nonlinear expectation possesses a nice dynamic feature, and it allows us to use many stochastic analysis tools. Replacing \(\mathbb{E}\) by \(\mathbb{E}_g\) in (1.2), we obtain a very natural extension of a classical stochastic optimal control problem, which can possibly take into account (subjective) nonlinear preferences. It turns out that \(g\)-expectation is closely related to the stochastic differential utility which was introduced by Duffie and Epstein [10] and further extended in [18], [2]. See [5], [6], [4] for some relevant information.

Now, let us make the above precise. Let \(g : [0, T] \times \mathbb{R} \times \mathbb{R} \times \Omega \to \mathbb{R}\) be a given map. Let \(p > 1\). For any \(\eta \in L^p_{\mathcal{F}_T}(\Omega; \mathbb{R})\), the set of all \(\mathcal{F}_T\)-measurable random variables such that \(\mathbb{E}|\eta|^p < \infty\), consider the following BSDE:

\[
\begin{align*}
\frac{dY(t)}{dt} &= g(t, Y(t), Z(t))dt + Z(t)dW(t), \quad t \in [0, T], \\
Y(T) &= \eta.
\end{align*}
\]

It is known that under some mild conditions, (1.4) admits a unique adapted solution \((Y(\cdot), Z(\cdot)) \equiv (Y(\cdot; \eta), Z(\cdot; \eta))\) [11]. According to [10], [18] (see also [28], [29]), we refer to \(Y(\cdot)\) as the stochastic differential utility process of \(\eta\) with the so-called...
aggregator \( g(\cdot) \). Then, we define

\[
\mathcal{E}_g[\eta] = Y(0; \eta) \quad \forall \eta \in L^p_T(\Omega; \mathbb{R}).
\]

When

\[
g(t, y, 0) = 0 \quad \forall (t, y) \in [0, T] \times \mathbb{R}, \ a.s.
\]

holds, \( \mathcal{E}_g[\eta] \) is called the \( g \)-expectation of \( \eta \) \[26\]; in this case, operator \( \mathcal{E}_g : L^p_T(\Omega) \to \mathbb{R} \) has all the major properties that \( \mathbb{E} \) has, except for the linearity. Based on this, let us replace (1.2) by the following:

\[
J_g(X_0, v(\cdot)) = \mathcal{E}_g \left\{ \int_0^T G(t, X(t), v(t))dt + F(X_0, X(T)) \right\} = Y(0),
\]

with \((Y(\cdot), Z(\cdot))\) being the unique adapted solution to the following BSDE:

\[
\begin{aligned}
dY(t) &= g(t, Y(t), Z(t))dt + Z(t)dW(t), \quad t \in [0, T], \\
Y(T) &= \int_0^T G(t, X(t), v(t))dt + F(X_0, X(T)).
\end{aligned}
\]

If we define

\[
X^0(t) = \int_0^t G(s, X(s), v(s))ds, \quad t \in [0, T],
\]

then the original problem becomes to minimize

\[
J_g(X_0, v(\cdot)) = Y(0) = \mathcal{E}_g \left\{ X^0(T) + F(X_0, X(T)) \right\}
\]

subject to the state equation

\[
\begin{aligned}
d \begin{pmatrix} X^0(t) \\ X(t) \end{pmatrix} &= \begin{pmatrix} G(t, X(t), v(t)) \\ B(t, X(t), v(t)) \end{pmatrix} dt + \begin{pmatrix} 0 \\ \Sigma(t, X(t), v(t)) \end{pmatrix} dW(t), \\
dY(t) &= g(t, Y(t), Z(t))dt + Z(t)dW(t), \\
X^0(0) &= \begin{pmatrix} 0 \\ X_0 \end{pmatrix}, \quad Y(T) = X^0(T) + F(X_0, X(T)).
\end{aligned}
\]

Equation (1.11) is a controlled forward-backward stochastic differential equation (FBSDE) with mixed initial-terminal conditions (see \[21\] for a general theory of classical FBSDEs, and see \[39\] for FBSDEs with mixed initial-terminal conditions). Thus we end up with an optimal control problem for an FBSDE, with state process \((X^0(\cdot), X(\cdot), Y(\cdot), Z(\cdot))\) and control process \(v(\cdot)\). It is not hard to imagine that the stochastic differential utility process \(Y(\cdot)\) could directly affect the process \(X(\cdot)\) through the drift/diffusion of the equation satisfied by \(X(\cdot)\). When this happens, we will end up with a controlled coupled FBSDE. One may further introduce the so-called forward-backward stochastic differential utility \[2\], which will directly lead to controlled coupled FBSDEs.

Our next motivation is the problem of leader-follower stochastic differential games \[36\] and/or the so-called principal-agent problem \[31\], \[7\]. Let us briefly look at such a situation. There are two players controlling the system

\[
\begin{aligned}
dX(t) &= b(t, X(t), u_1(t), u_2(t))dt \\
&\quad + \sigma(t, X(t), u_1(t), u_2(t))dW(t), \quad t \in [0, T], \\
X(0) &= X_0.
\end{aligned}
\]
We assume that the leader (or the principal) takes control \( u_1(\cdot) \) and that the follower (or the agent) takes control \( u_2(\cdot) \). Define
\[
(1.13) \quad J_i(u_1(\cdot), u_2(\cdot)) = \mathbb{E}\left[ \int_0^T f_i(t, X(t), u_1(t), u_2(t))dt + h_i(X(T)) \right], \quad i = 1, 2.
\]
Here, \( J_1(u_1(\cdot), u_2(\cdot)) \) and \( J_2(u_1(\cdot), u_2(\cdot)) \) are cost functionals of the leader and the follower, respectively. For any control \( u_1(\cdot) \) of the leader, the follower would like to choose a control \( \tilde{u}_2(\cdot) \) such that \( u_2(\cdot) \mapsto J_2(u_1(\cdot), u_2(\cdot)) \) is minimized. Suppose, under certain conditions, such an optimal control exists, which should be of the form \( \tilde{u}_2(\cdot) = \varphi(u_1(\cdot)) \) for some \( \varphi(\cdot) \). One can expect that the explicit form of \( \varphi(\cdot) \) will be very complicated. However, by deriving Pontryagin’s maximum principle for the optimal control \( \tilde{u}_2(\cdot) \), one could represent \( \varphi(\cdot) \) through the adapted solution to a coupled FBSDE [37], [40]. Then the leader faces a controlled FBSDE for which he/she would like to choose a control \( \tilde{u}_1(\cdot) \) so that \( u_1(\cdot) \mapsto J_1(u_1(\cdot), \varphi(u_1(\cdot))) \) is minimized. For a linear quadratic case, such an FBSDE was derived in [36], and for a general nonlinear case, see some discussions in [37]. In the principal-agent problem, the problem for the principal is essentially an optimal control problem for FBSDEs (see [31], [7]).

Our third motivation comes from an optimal investment problem in a financial market in which there are stocks (whose prices follow usual SDEs), bonds (whose prices follow ODEs with possibly random coefficients), and some European-type contingent claims (whose prices follow some BSDEs). In addition, there are some SDEs driving certain financial factors (such as interest rate, etc.). In such a setting, one has a controlled FBSDE.

Some studies relevant to the current paper can be found in [25], [34], [32], [31], [9], [12], [30], [17], [38], [33], and so on.

Motivated by the above, in this paper we are going to study the optimal control problem of the following fully coupled controlled FBSDE having mixed initial-terminal conditions:
\[
(1.14) \quad \begin{cases} 
    dx(t) = b(t, x(t), y(t), z(t), u(t))dt + \sigma(t, x(t), y(t), z(t), u(t))dW(t), \\
    dy(t) = g(t, x(t), y(t), z(t), u(t))dt + z(t)dw(t), \\
    x(0) = \gamma(x(T), y(0)), \\
    y(T) = h(x(T), y(0)), 
\end{cases}
\]
with the cost functional given by
\[
(1.15) \quad J(u(\cdot)) = \mathbb{E}\left[ f(x(T), y(0)) \right],
\]
where \( b(\cdot), \sigma(\cdot), g(\cdot), \gamma(\cdot), h(\cdot), f(\cdot) \) are given maps. In (1.14), the state process is any adapted solution \((x(\cdot), y(\cdot), z(\cdot))\) and the control process is \( u(\cdot) \). The main features of our problem are the following: (i) The state equation is a coupled FBSDE with mixed initial-terminal conditions; (ii) the control process \( u(\cdot) \) and the process \( z(\cdot) \) both appear not only in the drift terms \( b(\cdot) \) and \( g(\cdot) \) but also in the diffusion term \( \sigma(\cdot) \); and (iii) the control domain \( U \) is not assumed to be convex. Clearly, problems studied in [24], [25], [34], [32], [9], [17], [30], and [33] are all special cases of ours.

For any fixed control process \( u(\cdot) \), (1.14) is an FBSDE with mixed initial-terminal conditions and with random coefficients. In a recent work [39], this author introduced the notion of the Lyapunov operator which gives a certain compatibility among the coefficients of the FBSDE. When the existence of a Lyapunov operator is assumed, one
can obtain the well-posedness of the corresponding FBSDE. Further, under certain reasonable conditions, Lyapunov operators were constructed in [39], which includes the well-known monotone cases studied in [15], [35], [27], as well as some other more general cases. It is obvious that we may closely follow [39] to introduce Lyapunov operators for our controlled FBSDE (1.14) so that one has the well-posedness of the state equation for any given control process $u(\cdot)$. However, our main goal in this paper is to derive necessary conditions for optimal controls of our problem; the well-posedness of the state equation is not the main theme here. Hence, we will impose only the well-posedness of the state equation (1.14) as an assumption to avoid some technicalities not closely related to our main results.

Let us make some further remarks on (1.14)–(1.15). In (1.14), the dependence of $y(T)$ on $x(T)$ is classical (in the theory of FBSDEs [21]), and the dependence of $y(T)$ on $y(0)$ will allow us to transform a more general cost functional into a simpler form (see below). To explain the dependence of $x(0)$ on $(x(T), y(0))$, we note that if $y(\cdot)$ is regarded as a stochastic differential utility process, then the dependence of $x(0)$ on $y(0)$ means that the initial state $x(0)$ depends on the initial value $y(0)$ of the stochastic differential utility process $y(\cdot)$, and the dependence of $x(0)$ on $x(T)$ specifies some constraints on $(x(0), x(T))$ jointly. In the cost functional (1.15), the dependence on $x(T)$ appears in classical optimal control problems of Mayer type, which usually represents certain terminal rewarding/penalty, etc., and the dependence on $y(0)$ comes from the context of stochastic differential utility (see (1.10)).

Next, suppose that for the FBSDE state equation (1.14) one would like to further replace cost functional (1.15) by the following form which looks much more general:

\begin{equation}
J(u(\cdot)) = \mathbb{E}e^\gamma \left\{ \int_0^T b^0(t, x(t), y(t), z(t), u(t)) dt + h^0(x(T), y(0)) \right\} = y^0(0),
\end{equation}

with $(y^0(\cdot), z^0(\cdot))$ being the adapted solution of the following BSDE:

\begin{equation}
\begin{cases}
dy^0(t) = g^0(t, y^0(t), z^0(t)) dt + z^0(t) dW(t), \\
y^0(T) = x^0(T) + h^0(x(T), y(0))
\end{cases}
\end{equation}

for some map $g^0$, where

\begin{equation}
x^0(t) = \int_0^t b^0(s, x(s), y(s), z(s), u(s)) ds.
\end{equation}

Then we can augment the original state equation (1.14) as follows:

\begin{equation}
\begin{cases}
d\begin{pmatrix} x^0(t) \\ x(t) \end{pmatrix} = \begin{pmatrix} b^0(t, x(t), y(t), z(t), u(t)) \\ b(t, x(t), y(t), z(t), u(t)) \end{pmatrix} dt \\
\phantom{\begin{pmatrix} x^0(t) \\ x(t) \end{pmatrix} = } + \begin{pmatrix} 0 \\ \sigma(t, x(t), y(t), z(t), u(t)) \end{pmatrix} dW(t), \\
d\begin{pmatrix} y^0(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} g^0(t, y^0(t), z^0(t)) \\ g(t, x(t), y(t), z(t), u(t)) \end{pmatrix} dt + \begin{pmatrix} z^0(t) \\ z(t) \end{pmatrix} dW(t), \\
x^0(0) = \begin{pmatrix} 0 \\ \gamma(x(T), y(0)) \end{pmatrix}, \\
y^0(T) = \begin{pmatrix} x^0(T) + h^0(x(T), y(0)) \\ h(x(T), y(0)) \end{pmatrix}
\end{cases}
\end{equation}
which is still a form of (1.14) with two dimensions higher. The cost functional (1.16)
becomes a form of (1.15). Hence, state equation (1.14) with cost functional (1.15) is
already general enough. Note that as long as either \( h^0 \) or \( h \) depends on \( y(0) \), we have
the dependence of \( y^0(T) \) on \( y(0) \) in the above, which gives one reason why we should
assume the dependence of \( y(T) \) on \( y(0) \) in (1.14).

The rest of this paper is organized as follows. In section 2, we make some prepara-
tions; among other things, a major preparation is to make a reduction of our optimal
total problem for FBSDEs with mixed initial-terminal conditions. Section 3 is de-

toted to the statement of our main results, together with some discussions on special
cases. In section 4, a proof of our optimality variational principle will be presented. A
linear-quadratic situation will be briefly discussed in section 5, which serves as an illus-
trative example. More consequences of this paper will be presented in a forthcoming
publication.

2. Preliminaries; problem reduction. In this section, we will make some preliminaries.

Let \( T > 0 \) be fixed, let \( U \) be a separable metric space, and denote \( M = \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \). Any generic point in \( M \) is denoted by \( \theta \equiv (x, y, z) \). Let \( U[0, T] \) be the set
of all \( \mathcal{F} \)-progressively measurable processes \( u : [0, T] \times \Omega \to U \). Any \( u(\cdot) \in U[0, T] \) is
referred to as an admissible control process. Now, we introduce some spaces. For any
Euclidean space \( H \) (with norm \( \| \cdot \| \)) and \( p, q \geq 1 \), let

\[
L^p_{\mathcal{F}}(\Omega; H) = \left\{ \xi : \Omega \to H \mid \xi \text{ is } \mathcal{F}_T \text{-measurable, } \mathbb{E}[|\xi|^p] < \infty \right\},
\]

\[
L^p(\Omega; L^q(0, T; H)) = \left\{ \varphi : [0, T] \times \Omega \to H \mid \varphi \text{ is } \mathcal{F}_t \text{-adapted, } \mathbb{E}\left( \int_0^T |\varphi(t)|^q dt \right)^{\frac{p}{q}} < \infty \right\},
\]

\[
L^p(\Omega; C([0, T]; H)) = \left\{ \varphi : [0, T] \times \Omega \to H \mid \varphi \text{ is continuous, } \mathcal{F}_t \text{-adapted, } \mathbb{E}\left( \sup_{t \in [0, T]} |\varphi(t)|^p \right) < \infty \right\}.
\]

We denote

\[
L^p_{\mathcal{F}}(0, T; H) = L^p_{\mathcal{F}}(\Omega; L^p(0, T; H)),
\]

\[
Z^p[0, T] = L^p_{\mathcal{F}}(\Omega; L^2(0, T; \mathbb{R}^m)), \quad X^p = L^p_{\mathcal{F}}(\Omega; \mathbb{R}^n),
\]

\[
\mathcal{M}^p[0, T] \equiv \Delta L^p_{\mathcal{F}}(\Omega; C([0, T]; \mathbb{R}^m)) \times L^p_{\mathcal{F}}(\Omega; C([0, T]; \mathbb{R}^m)) \times L^p_{\mathcal{F}}(\Omega; L^2(0, T; \mathbb{R}^m)).
\]

Clearly, \( \mathcal{M}^p[0, T] \) is a Banach space. Any process in \( \mathcal{M}^p[0, T] \) is denoted by
\( \Theta(\cdot) \equiv (x(\cdot), y(\cdot), z(\cdot)) \), whose norm is given by

\[
\|\Theta(\cdot)\|_{\mathcal{M}^p[0, T]} \equiv \Delta \left\{ \mathbb{E}\left[ \sup_{t \in [0, T]} |x(t)|^p + \sup_{t \in [0, T]} |y(t)|^p + \left( \int_0^T |z(t)|^2 dt \right)^{\frac{p}{2}} \right] \right\}^{\frac{1}{p}}.
\]

For any \( u(\cdot) \in U[0, T] \), by an adapted solution to (1.14) we mean a process
\( \Theta(\cdot) \equiv (x(\cdot), y(\cdot), z(\cdot)) \in \mathcal{M}^p[0, T] \) that satisfies (1.14) in the usual It\( \hat{o} \) sense (see [39]).
We now consider the following FBSDE:

\[
\begin{aligned}
    &dx(t) = b(t, x(t), y(t), z(t), u(t))dt + \sigma(t, x(t), y(t), z(t), u(t))dW(t), \\
    &dy(t) = g(t, x(t), y(t), z(t), u(t))dt + z(t)dW(t), \\
    &x(0) = \gamma(x(T), y(0)) + \gamma_0, \\
    &y(T) = h(x(T), y(0)) + h_T,
\end{aligned}
\]

(2.2)

with \((\gamma_0, h_T) \in \mathbb{R}^n \times \mathcal{A}_m^2\). Clearly, when \((\gamma_0, h_T) = 0\), (2.2) is reduced to (1.14). Note that for any admissible control \(u(\cdot) \in \mathcal{U}[0, T]\), (2.2) is a coupled FBSDE with mixed initial-terminal conditions, and with random coefficients. Before going further, let us specify the coefficients of (2.2) as follows:

\[
\begin{aligned}
    &b, \sigma : [0, T] \times \mathcal{M} \times U \times \Omega \rightarrow \mathbb{R}^n, \\
    &g : [0, T] \times \mathcal{M} \times U \rightarrow \mathbb{R}^m, \\
    &\gamma : \mathcal{A}_m^2 \times \mathbb{R}^m \rightarrow \mathbb{R}^n, \\
    &h : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \rightarrow \mathbb{R}^m.
\end{aligned}
\]

We introduce the following basic assumption which will be assumed throughout the paper.

(H0) For any \((\gamma_0, h_T) \in \mathbb{R}^n \times \mathcal{A}_m^2\) and \(u(\cdot) \in \mathcal{U}[0, T]\), FBSDE (2.2) admits a unique adapted solution \(\Theta(\cdot) \equiv (x(\cdot), y(\cdot), z(\cdot)) \in \mathcal{M}^2[0, T]\). Moreover, the following estimate holds:

\[
||\Theta(\cdot)||^2_{\mathcal{M}^2[0, T]} = \mathbb{E}\left\{ \sup_{t \in [0, T]} |x(t)|^2 + \sup_{t \in [0, T]} |y(t)|^2 + \int_0^T |z(t)|^2 dt \right\}
\]

(2.3)

\[
\leq K\mathbb{E}\left\{ |\gamma(0, 0) + \gamma_0|^2 + |h(0, 0) + h_T|^2 + \left( \int_0^T |b(t, 0, u(t))| dt \right)^2 \\
+ \left( \int_0^T |g(t, 0, u(t))| dt \right)^2 + \int_0^T |\sigma(t, 0, u(t))|^2 dt \right\}.
\]

Hereafter, \(K > 0\) stands for a generic constant which can be different at different places. Further, if \(\tilde{\Theta}(\cdot) \equiv (\tilde{x}(\cdot), \tilde{y}(\cdot), \tilde{z}(\cdot)) \in \mathcal{M}^2[0, T]\) is the unique adapted solution of (2.2) with \((\gamma_0, h_T) \in \mathbb{R}^n \times \mathcal{A}_m^2\) and \(\tilde{u}(\cdot) \in \mathcal{U}[0, T]\), respectively, then the following stability estimate holds:

\[
||\tilde{\Theta}(\cdot) - \Theta(\cdot)||^2_{\mathcal{M}^2[0, T]} = \mathbb{E}\left\{ \sup_{t \in [0, T]} |\tilde{x}(t) - x(t)|^2 + \sup_{t \in [0, T]} |\tilde{y}(t) - y(t)|^2 + \int_0^T |\tilde{z}(t) - z(t)|^2 dt \right\}
\]

(2.4)

\[
\leq K\mathbb{E}\left\{ |\tilde{\gamma}_0 - \gamma_0|^2 + |\tilde{h}_T - h_T|^2 + \left( \int_0^T |b(t, \Theta(t), \tilde{u}(t)) - b(t, \Theta(t), u(t))| dt \right)^2 \\
+ \left( \int_0^T |g(t, \Theta(t), \tilde{u}(t)) - g(t, \Theta(t), u(t))| dt \right)^2 + \int_0^T |\sigma(t, \Theta(t), \tilde{u}(t)) - \sigma(t, \Theta(t), u(t))|^2 dt \right\}.
\]

As we mentioned in the introduction, by adopting the ideas from [39], we know that under certain suitable conditions, the Lyapunov operators will exist for (2.2), which leads to hypothesis (H0); in particular, this will be the case if one assumes certain monotonicity conditions for the coefficients. We should point out that (1.14) and (2.2) have the same Lyapunov operators. Hence, the well-posedness of (2.2) via
Our main goal in this paper is to derive a set of necessary conditions for the optimal controls of Problem (C). Following the standard idea of deriving necessary conditions for optimal control processes, due to the fact that $U$ is not assumed to be convex one needs to use spike variation for the control process, and then to try obtaining a Taylor-type expansion for the state process and the cost functional with respect to the spike variation of the control process, followed by some suitable duality relations to get a maximum principle of Pontryagin type. With such a direct and natural approach, one will face two major difficulties: (i) The derivation of Taylor expansion of the state process $(x(\cdot), y(\cdot), z(\cdot))$ with respect to the spike variation of the control process $u(\cdot)$ is technically difficult (even for the classical case of separated initial and terminal conditions: $x(0) = x_0$ and $y(T) = h(x(T))$). The main reasons are that both $z(\cdot)$ and $u(\cdot)$ appear in the diffusion coefficient of the forward equation and that the regularity/integrability of process $z(\cdot)$ (as a part of the state process) seems to be not enough in the case when a second order expansion is necessary. Note that in [25], [34], [32], [9], [30], and [17], due to the special structure of the problems, the second order expansion is not necessary when one tries to derive the necessary conditions for optimal controls (see more about this in section 3). (ii) Due to the mixed initial-terminal conditions in the state equation, even if we have well-posedness of the state equation via the Lyapunov operator introduced in [39], which automatically also gives the well-posedness of the linearized state equation, the well-posedness of the first adjoint equation seems to be not guaranteed. Moreover, due to the forward-backward nature of the state equation, it seems to be unexpectedly complicated just for formally deriving the second adjoint equation, without even touching the well-posedness of such an equation!

Due to the above item (i), to the best of our knowledge, no essential progress has been made in deriving a maximum principle for optimal controls of fully coupled FBSDEs since the case with $U$ being convex and the initial-terminal conditions being separated was studied more than 10 years ago in [32]. Note that in [30], although the control domain $U$ was not assumed to be convex, the control process was not allowed to appear in the forward diffusion coefficient $\sigma(\cdot)$, whereas, in [33], the process $z(\cdot)$ is not allowed to appear in $\sigma(\cdot)$. Further, by item (ii), the above-mentioned direct approach seems to be even hopeless. Therefore, a different approach is really needed in order to make some essential progress. A main contribution of the current paper is the discovery of the following reduction of Problem (C), inspired by the study of an optimal control problem for infinite-dimensional evolution equations with mixed initial-terminal constraints [19], [20]. Such an approach leads to an essentially clean sweep of a research corner of optimal control for FBSDEs. Let us make it precise now.
Consider the following controlled variational initial value problem for a system of SDEs:

\[
\begin{aligned}
\begin{cases}
    dx(t) = b(t, x(t), y(t), z(t), u(t))dt + \sigma(t, x(t), y(t), z(t), u(t))dW(t), \\
    dy(t) = g(t, x(t), y(t), z(t), u(t))dt + z(t)dB(t), \\
    x(0) = x_0, & y(0) = y_0,
\end{cases}
\end{aligned}
\tag{2.6}
\]

where \((x(\cdot), y(\cdot))\) is regarded as the state process and \((z(\cdot), u(\cdot))\) is regarded as the control process. We introduce the following hypothesis.

\textbf{(H1)} Maps \(b, \sigma : [0, T] \times \mathbb{M} \times U \times \Omega \to \mathbb{R}^n\), \(g : [0, T] \times \mathbb{M} \times U \to \mathbb{R}^m\) satisfy the following conditions: (i) For any \((\theta, u) \in \mathbb{M} \times U\), \((b(t, \theta, u), \sigma(t, \theta, u), g(t, \theta, u))\) is \(\mathbb{F}\)-progressively measurable; (ii) there exists a constant \(L > 0\) such that

\[
\begin{aligned}
&|b(t, x, y, z, u) - b(t, \bar{x}, \bar{y}, z, u)| + |\sigma(t, x, y, z, u) - \sigma(t, \bar{x}, \bar{y}, z, u)| \\
&+ |g(t, x, y, z, u) - g(t, \bar{x}, \bar{y}, z, u)| \
&\leq L \left( |x - \bar{x}| + |y - \bar{y}| \right)
\end{aligned}
\tag{2.7}
\]

\(\forall (t, z, u) \in [0, T] \times \mathbb{R}^m \times U\), \((x, y)\), \((\bar{x}, \bar{y})\) \(\in \mathbb{R}^n \times \mathbb{R}^m\), a.s.

and

\[
\begin{aligned}
&|b(\cdot, 0, 0, z(\cdot), u(\cdot))| + |\sigma(\cdot, 0, 0, z(\cdot), u(\cdot))|^2 \\
&+ |g(\cdot, 0, 0, z(\cdot), u(\cdot))| \in L^2_b(0, T; \mathbb{R})
\end{aligned}
\tag{2.8}
\]

\(\forall z(\cdot) \in \mathcal{Z}[0, T]\), \(u(\cdot) \in \mathcal{U}[0, T]\).

It is standard that, under (H1), for any \((z(\cdot), u(\cdot)) \in \mathcal{Z}^2[0, T] \times \mathcal{U}[0, T]\) and \((x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m\), there exists a unique strong solution to (2.6),

\[
(x(\cdot), y(\cdot)) \equiv \left( x(\cdot), x_0, y_0, z(\cdot), u(\cdot) \right)
\in L^2_\mathbb{F}(\Omega, C([0, T]; \mathbb{R}^n)) \times L^2_\mathbb{F}(\Omega, C([0, T]; \mathbb{R}^m)),
\]

depending on the 4-tuple \((x_0, y_0, z(\cdot), u(\cdot))\). Next, we regard the original mixed initial-terminal conditions as mixed initial-terminal state constraints:

\[
(2.9) \quad x_0 = \gamma(x(T), y_0), \quad y(T) = h(x(T), y_0).
\]

Let \(\mathcal{A}\) be the set of all 4-tuples \((x_0, y_0, z(\cdot), u(\cdot))\) \(\in \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{Z}^2[0, T] \times \mathcal{U}[0, T]\) such that the unique corresponding state process \((x(\cdot), y(\cdot))\) satisfies the constraints (2.9). Note that hypothesis (H0), the well-posedness of the FBSDE with mixed initial-terminal conditions, implies that for any \(u(\cdot) \in \mathcal{U}[0, T]\), there exists a unique triple \((x_0, y_0, z(\cdot))\) \(\in \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{Z}^2[0, T]\) such that state equation (2.6) admits a unique state process \((x(\cdot), y(\cdot))\) \(\in L^2_\mathbb{F}(\Omega, C([0, T]; \mathbb{R}^n)) \times L^2_\mathbb{F}(\Omega, C([0, T]; \mathbb{R}^m))\) satisfying the mixed initial-terminal state constraints (2.9). Hence, (H0) and (H1) imply \(\mathcal{A} \neq \emptyset\).

We rewrite the cost functional as follows:

\[
(2.10) \quad J(x_0, y_0, z(\cdot), u(\cdot)) = \mathbb{E} \left[ f(x(T), y_0) \right].
\]

Now, we can pose the following problem.

\textbf{Problem (C).} Find \((\bar{x}_0, \bar{y}_0, \bar{z}(\cdot), \bar{u}(\cdot)) \in \mathcal{A}\) such that

\[
(2.11) \quad J(\bar{x}_0, \bar{y}_0, \bar{z}(\cdot), \bar{u}(\cdot)) = \inf_{(x_0, y_0, z(\cdot), u(\cdot)) \in \mathcal{A}} J(x_0, y_0, z(\cdot), u(\cdot)).
\]
We respectively refer to \((\bar{x}(\cdot), \bar{y}(\cdot))\) as an optimal control process, to \((\bar{x}(\cdot), \bar{y}(\cdot))\) as the corresponding optimal state process, and to \((\bar{x}_0, \bar{y}_0, \bar{z}(\cdot), \bar{u}(\cdot))\) as an optimal 4-tuple, of Problem \((\hat{C})\). It is clear that Problems \((C)\) and \((\hat{C})\) are equivalent. The advantage of Problem \((\hat{C})\) is that one does not need much regularity/integrability on \(z(\cdot)\) since it is treated as a control process; the disadvantage is that one has to treat mixed initial-terminal state constraints \((2.9)\). We have to point out that the state constraints \((2.9)\) represent infinitely many equality constraints, since both \(y(T)\) and \(h(x(T), y_0)\) are in \(\mathcal{X}_m^2 \equiv L^2_{\mathcal{P}}(\Omega; \mathbb{R}^m)\) of infinite dimensions. Recall that for optimal control problems of deterministic infinite-dimensional evolution systems with mixed initial-terminal state constraints, in order to ensure the nontriviality of the maximum principle (which is guaranteed by the nontriviality of the solution to the adjoint equation, which is called the costate process), one has to assume some finite codimensionality condition for a set relevant to the mixed constraint and the reachable set of the linearized state equation \([19], [20]\). It turns out that under \((H0)\), we are able to introduce some abbreviation notations. First, we make the following convention: For \(d \in \mathbb{R}\), \(\lambda \in \mathbb{R}\) and \(\gamma \in \mathbb{R}\), let \(\hat{\lambda}(\cdot)\) denote the gradient of \(\lambda(\cdot)\) or \(\lambda_{\gamma}(\cdot)\) denote the gradient of \(\lambda_{\gamma}(\cdot)\). If the function \(\lambda(\cdot)\) is twice continuously differentiable, \(\gamma(\cdot)\) is twice differentiable with the (partial) derivatives up to order 2 being all uniformly bounded and uniformly Lipschitz continuous. The map \(f : \mathcal{X}_m^2 \times \mathbb{R}^m \to \mathcal{X}_m^2\) is twice (Fréchet) differentiable with the (partial or Fréchet) derivatives up to order 2 being all uniformly bounded and uniformly Lipschitz continuous: the map \(f : \mathcal{X}_m^2 \times \mathbb{R}^m \to \mathcal{X}_m^2\) is twice (Fréchet) differentiable with the (partial or Fréchet) derivatives up to order 2 being Lipschitz continuous.

Note that if we introduce the following state constraint:

\begin{equation}
L(x_0, y_0, x(T), y(T)) = E \left[ |x_0 - \gamma(x(T), y_0)|^2 + |y(T) - h(x(T), y_0)|^2 \right] = 0,
\end{equation}

then one still can formulate an equivalent optimal control problem. However, the resulting optimality variational principle will be trivial. This is mainly due to the fact that

\begin{equation}
DL(\bar{x}(0), \bar{y}(0), \bar{x}(T), \bar{y}(T)) = 0,
\end{equation}

where \(DL\) is the gradient of \(L\), and \((\bar{x}(\cdot), \bar{y}(\cdot))\) is an optimal state process.

3. Optimality variational principle and some special cases. In this section, we will state the set of necessary conditions for optimal controls of our Problem \((\hat{C})\), which we call the optimality variational principle for controlled FBSDE with mixed initial-terminal conditions. Also, some interesting special cases will be discussed.

Let us now introduce the following further assumptions.

\((H2)\) Maps \(\theta \mapsto (b(t, \theta, u), \sigma(t, \theta, u), g(t, \theta, u))\) are twice continuously differentiable, with the (partial) derivatives up to order 2 being uniformly bounded, Lipschitz continuous in \(\theta \in \mathbb{M}\), and continuous in \(u \in U\).

\((H3)\) The maps \(\gamma : \mathcal{X}_m^2 \times \mathbb{R}^m \to \mathbb{R}^n\), \(h : \mathcal{X}_m^2 \times \mathbb{R}^m \to \mathcal{X}_m^2\) are twice (Fréchet) differentiable with the (partial or Fréchet) derivatives up to order 2 being all uniformly bounded and uniformly Lipschitz continuous; the map \(f : \mathcal{X}_m^2 \times \mathbb{R}^m \to \mathcal{X}_m^2\) is twice (Fréchet) differentiable with the (partial or Fréchet) derivatives up to order 2 being Lipschitz continuous.

Note that \((H2)\) and \((H3)\) are more than enough, and some of the conditions can be substantially relaxed. But we prefer not to get into those details for the simplicity of our presentation. We note that the conditions imposed on \(f(\cdot)\) allow it to be a quadratic form in \((x_T, y_0) \in \mathcal{X}_m^2 \times \mathbb{R}^m\). Next, to simplify our presentation, we now introduce some abbreviation notations. First, we make the following convention: For
any differentiable map \( a \equiv (a^1, \ldots, a^r)^T: \mathbb{R}^n \to \mathbb{R}^r \), let
\[
a_x = \begin{pmatrix}
a^1_{x_1} & a^1_{x_2} & \cdots & a^1_{x_n} \\
a^2_{x_1} & a^2_{x_2} & \cdots & a^2_{x_n} \\
\vdots & \vdots & & \vdots \\
a^r_{x_1} & a^r_{x_2} & \cdots & a^r_{x_n}
\end{pmatrix}: \mathbb{R}^n \to \mathbb{R}^{r \times n}.
\]

In particular, for \( \ell = 1 \), \( a_x = (a_{x_1}, \ldots, a_{x_n}) \in \mathbb{R}^{1 \times n} \) is an \( n \)-dimensional row vector. Also, for any twice differentiable function \( a: \mathbb{R}^n \to \mathbb{R} \), the Hessian is given by
\[
a_{xx} \equiv (a_x^T)_x = \begin{pmatrix}
a_{x_1 x_1} & a_{x_1 x_2} & \cdots & a_{x_1 x_n} \\
a_{x_2 x_1} & a_{x_2 x_2} & \cdots & a_{x_2 x_n} \\
\vdots & \vdots & & \vdots \\
a_{x_n x_1} & a_{x_n x_2} & \cdots & a_{x_n x_n}
\end{pmatrix}: \mathbb{R}^n \to \mathbb{S}^n;
\]
hereafter, \( \mathbb{S}^n \) stands for the set of all real \((n \times n)\) symmetric matrices. On the other hand, for a twice differentiable function \( a: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \) (denoting \((x, y) \in \mathbb{R}^n \times \mathbb{R}^m\)), we denote
\[
a_{xy} \equiv (a_x^T)_y = \begin{pmatrix}
a_{x_1 y_1} & a_{x_1 y_2} & \cdots & a_{x_1 y_m} \\
a_{x_2 y_1} & a_{x_2 y_2} & \cdots & a_{x_2 y_m} \\
\vdots & \vdots & & \vdots \\
a_{x_n y_1} & a_{x_n y_2} & \cdots & a_{x_n y_m}
\end{pmatrix}: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n \times m}.
\]

Now, let \((\bar{x}(\cdot), \bar{y}(\cdot), \bar{z}(\cdot), \bar{u}(\cdot))\) be an optimal 4-tuple of Problem (C). For \( a(\cdot) = b(\cdot), g(\cdot), \sigma(\cdot) \), we denote
\[
\begin{align*}
\bar{a}_x(t) &= a_x(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)), \\
\bar{a}_y(t) &= a_y(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)), \\
\bar{a}_z(t) &= a_z(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)),
\end{align*}
\]
and let
\[
\bar{B}_x(t) = \begin{pmatrix}
\bar{b}_x(t) \\
\bar{b}_y(t) \\
g_y(t)
\end{pmatrix}, \quad \Sigma_x(t) = \begin{pmatrix}
\bar{\sigma}_x(t) & \bar{\sigma}_y(t) \\
0 & 0
\end{pmatrix}.
\]

Also, for \( a(\cdot) = b^{i}(\cdot), \sigma^{i}(\cdot), g^{i}(\cdot) \), \( 1 \leq i \leq n, 1 \leq j \leq m \), let
\[
\begin{align*}
\bar{a}_{i x}(t) &= a_{i x}(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)), \\
\bar{a}_{i y}(t) &= a_{i y}(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)), \\
\bar{a}_{i y}(t) &= a_{i y}(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)), \\
\bar{a}_{i z}(t) &= a_{i z}(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)),
\end{align*}
\]
and for \( \rho = f^i, h^j \), we denote
\[
\begin{align*}
\bar{p}_y(t) &= \rho_y(t, \bar{x}(T), \bar{y}(0)), \\
\bar{p}_{x T} &= \rho_{x T}(\bar{x}(T), \bar{y}(0)), \\
\bar{p}_{y y} &= \rho_{y y}(\bar{x}(T), \bar{y}(0)), \\
\bar{p}_{x x y} &= \rho_{x x y}(\bar{x}(T), \bar{y}(0)), \\
\bar{p}_{x y} &= \rho_{x y}(\bar{x}(T), \bar{y}(0)).
\end{align*}
\]

Our main result is the following optimality variational principle of Problem (C).
Theorem 3.1. Let (H0)–(H3) hold, and let $(\bar{x}(\cdot), \bar{y}(\cdot), \bar{z}(\cdot), \bar{u}(\cdot)) \in \mathcal{A}$ be an optimal 4-tuple of Problem (C). Then there exists an adapted solution $(\xi(\cdot), \eta(\cdot), \zeta(\cdot)) \in L^2(\Omega; C([0, T]; \mathbb{R}^m)) \times L^2(\Omega; C([0, T]; \mathbb{R}^n)) \times L^2(0, T; \mathbb{R}^n)$ to the following FBSDE:

\[
\begin{aligned}
\left\{ \begin{array}{l}
\, d\xi(t) = -\left[ \bar{g}_y(t)\xi(t) + \bar{b}_y(t)\eta(t) + \bar{\sigma}_y(t)\zeta(t) \right] dt \\
-\left[ \bar{g}_x(t)\xi(t) + \bar{b}_x(t)\eta(t) + \bar{\sigma}_x(t)\zeta(t) \right] dW(t),
\end{array} \right.
\end{aligned}
\tag{3.1}
\]

\[
\begin{aligned}
\, d\eta(t) = -\left[ \bar{g}_x(t)\xi(t) + \bar{b}_x(t)\eta(t) + \bar{\sigma}_x(t)\zeta(t) \right] dt + \zeta(t) dW(t),
\end{aligned}
\]

\[
\xi(0) = \mathbb{E}\left[ \bar{h}_y(T)\xi(T) - \bar{\gamma}_y(0) + \mathbb{E}\bar{f}_y \right],
\]

\[
\eta(T) = -\bar{h}_x(T)\xi(T) + \bar{\gamma}_x(0) + \bar{f}_x.
\]

Let $(\Phi(\cdot), \Psi(\cdot)) \in L^2(\Omega; C([0, T]; \mathbb{R}^{m+n})) \times L^2(0, T; \mathbb{R}^{m+n})$ be the unique adapted solution to the following BSDE:

\[
\begin{aligned}
\left\{ \begin{array}{l}
\, d\Phi(t) = -\left[ \bar{B}_X(t)^\top \Phi(t) + \bar{\Sigma}_X(t)^\top \Psi(t) \right] dt + \Psi(t) dW(t),
\end{array} \right.
\end{aligned}
\tag{3.2}
\]

\[
\Phi(T) = -\left( \sum_{j=1}^m \xi_j(T)\bar{h}_x + \sum_{i=1}^n \eta_i(0)\bar{\gamma}_x + \bar{f}_x \right),
\]

and let $(P(\cdot), Q(\cdot)) \in L^2(\Omega; C([0, T]; \mathbb{S}^{n+m})) \times L^2(0, T; \mathbb{S}^{n+m})$ be the unique adapted solution to the following BSDE:

\[
\begin{aligned}
\left\{ \begin{array}{l}
\, dP(t) = -\left[ \bar{B}_X(t)^\top P(t) + P(t)\bar{B}_X(t) + \bar{\Sigma}_X(t)^\top P(t)\bar{\Sigma}_X(t) \\
+ \bar{\Sigma}_X(t)^\top Q(t) + Q(t)\bar{\Sigma}_X(t) + \bar{H}_{XX}(t) \right] dt + Q(t) dW(t),
\end{array} \right.
\end{aligned}
\tag{3.3}
\]

\[
\begin{aligned}
P(T) = -\left( \sum_{j=1}^m \xi_j(T)\bar{h}_x + \sum_{i=1}^n \eta_i(0)\bar{\gamma}_x + \bar{f}_x \right),
\end{aligned}
\]

where $\bar{H}_{XX}(t)$, taking values in $\mathbb{S}^{n+m}$, is defined by the following:

\[
\bar{H}_{XX}(t) = \sum_{j=1}^m \begin{pmatrix}
\xi_j(t)\bar{g}_{xx}(t) & \xi_j(t)\bar{g}_{xy}(t) \\
\xi_j(t)\bar{g}_{yx}(t) & \xi_j(t)\bar{g}_{yy}(t)
\end{pmatrix}
\tag{3.4}
\]

\[
+ \sum_{i=1}^n \begin{pmatrix}
\eta_i(t)\bar{b}_{xx}(t) + \zeta_i(t)\bar{\sigma}_{xx}(t) & \eta_i(t)\bar{b}_{xy}(t) + \zeta_i(t)\bar{\sigma}_{xy}(t) \\
\eta_i(t)\bar{b}_{yx}(t) + \zeta_i(t)\bar{\sigma}_{yx}(t) & \eta_i(t)\bar{b}_{yy}(t) + \zeta_i(t)\bar{\sigma}_{yy}(t)
\end{pmatrix}.
\]

Then

\[
\begin{pmatrix}
0 \\
\mathbb{E}\left[ \sum_{j=1}^m \xi_j(T)\bar{h}_y \right] - \sum_{i=1}^n \eta_i(0)\bar{\gamma}_y + \mathbb{E}\bar{f}_y
\end{pmatrix} 
\geq 0,
\]
\[
\sum_{j=1}^{m} \xi_j(t) \bar{g}_j(t) + \sum_{i=1}^{n} \left[ \eta_i(t) \bar{b}_i(t) + \zeta_i(t) \bar{\sigma}_i(t) \right] \\
- \left( \begin{array}{c} \bar{\sigma}_z(t) \\ I \end{array} \right)^T P(t) \begin{pmatrix} \bar{\sigma}_z(t) \\ I \end{pmatrix} \geq 0 \quad \text{a.e. } t \in [0,T], \text{ a.s.}
\]

and

\[
H(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), u, \xi(t), \eta(t), \zeta(t)) \\
- H(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t), \xi(t), \eta(t), \zeta(t)) \\
- \frac{1}{2} \left[ P(t) \left[ \sigma(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), u) - \sigma(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)) \right] \right] \\
\sigma(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), u) - \sigma(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)) \geq 0, \\
\forall u \in U \quad \text{a.e. } t \in [0,T], \text{ a.s. } \omega \in \Omega,
\]

with \( P(\cdot) = \left( \begin{array}{cc} P_1(\cdot) \\ P_2(\cdot) \end{array} \right) \), and \( H(t, x, y, z, u, \xi, \eta, \zeta) \) being the Hamiltonian of Problem (C) defined as follows:

\[
H(t, x, y, z, u, \xi, \eta, \zeta) \\
= \langle \xi, g(t, x, y, z, u) \rangle + \langle \eta, b(t, x, y, z, u) \rangle + \langle \zeta, \sigma(t, x, y, z, u) \rangle, \\
(t, u) \in [0,T] \times U, \ x, \eta, \zeta \in \mathbb{R}^n, \ y, z, \xi \in \mathbb{R}^m, \text{ a.s.}
\]

In the above, we refer to (3.1) and (3.3) as the first and second adjoint equations, respectively; we refer to (3.7) as the optimality variational inequality.\(^1\)

Let us make some remarks.

- In the case that the dependence of \( \gamma, h, f \) on \( x_T \) and \( y_0 \) is separated, or, equivalently,

\[
\gamma_{x_T y_0}(x_T, y_0) = 0, \quad h_{x_T y_0}(x_T, y_0) = 0, \quad f_{x_T y_0}(x_T, y_0) = 0,
\]

then \( (\Phi(\cdot), \Psi(\cdot)) = 0 \).

- In the mixed initial-terminal conditions of the first adjoint equation (3.1), as long as \( \gamma \) depends on \( x_T, \eta(T) \) depends on \( \eta(0) \). Comparing (3.1) and (1.14), we see another reason why in (1.14) it is natural to have the dependence of \( y(T) \) on \( y(0) \).

- In the classical stochastic optimal control case, the first adjoint equation is a linear BSDE whose well-posedness can be established without optimal control theory. However, in the current case, the first adjoint equation (3.1) is a linear FBSDE with mixed initial-terminal conditions. The well-posedness of such an equation is not guaranteed by (H0), which only guarantees the well-posedness of the state equation (1.14). The above theorem tells us that the solvability of (3.1) is actually due to the existence of optimal 4-tuples.

Let us now look at some special cases. It is interesting that even for those special cases, our theorem not only recovers the known results, but also provides some additional necessary conditions.

1. A classical stochastic optimal control problem.

\(^1\)It does not seem very proper to call (3.7) either a maximum or a minimum condition.
Consider a controlled SDE

$$\begin{align*}
&(3.9) \quad \begin{cases}
dx(t) = b(t, x(t), u(t))dt + \sigma(t, x(t), u(t))dW(t), & t \in [0, T], \\
x(0) = \gamma_0,
\end{cases} \\
\end{align*}$$

with cost functional

$$\begin{align*}
&(3.10) \quad J(u(\cdot)) = \mathbb{E} \left[ f(x(T)) \right].
\end{align*}$$

In this case, we have

$$\begin{align*}
&(\dot{b}(t, x, y, z, u) = b(t, x, u), \quad \sigma(t, x, y, z, u) = \sigma(t, u), \quad g(t, x, y, z, u) = 0, \\
&\gamma(x_T, y_0) = \gamma_0, \quad h(x_T, y_0) = 0, \quad f(x_T, y_0) = f(x_T).
\end{align*}$$

Hence, we have, by some direct computation/observation,

$$\begin{align*}
&(3.12) \quad \begin{cases}
y(\cdot) = z(\cdot) = 0, \quad \xi(\cdot) = 0, \quad \Phi(\cdot) = \Psi(\cdot) = 0, \\
P_2(\cdot) = Q_2(\cdot) = 0, \quad P_3(\cdot) = Q_3(\cdot) = 0.
\end{cases}
\end{align*}$$

Consequently, (3.1) and (3.3) become

$$\begin{align*}
&(3.11) \quad \begin{cases}
d\eta(t) = - \left[ \bar{g}_x(t)^\top \xi(t) + \bar{b}_x(t)^\top \eta(t) + \bar{\sigma}_x(t)^\top \zeta(t) \right] dt \\
\quad + \zeta(t) dW(t), & t \in [0, T], \\
\eta(T) = - \bar{f}_T^\top
\end{cases}
\end{align*}$$

and

$$\begin{align*}
&(3.12) \quad \begin{cases}
dP_1(t) = - \left[ \bar{b}_x(t)^\top P_1(t) + P_1(t) \bar{b}_x(t) + \bar{\sigma}_x(t)^\top P_1(t) \bar{\sigma}_x(t) \\
\quad + \bar{\sigma}_x(t)^\top Q_1(t) + Q_1(t) \bar{\sigma}_x(t) + \sum_{j=1}^m \xi_j(t) \bar{g}_x(t)
\quad + \sum_{i=1}^n \left( \eta_i(t) \bar{b}_{xx}(t) + \zeta_i(t) \bar{\sigma}_{xx}(t) \right) \right] dt \\
\quad + Q_1(t) dW(t), & t \in [0, T], \\
P_1(T) = - \bar{f}_{xx}.
\end{cases}
\end{align*}$$

The optimality variational inequality (3.7) is reduced to the following:

$$\begin{align*}
&H(t, \bar{x}(t), u, \eta(t), \zeta(t)) - H(t, \bar{x}(t), \bar{u}(t), \eta(t), \zeta(t)) \\
&\quad - \frac{1}{2} \left( P_1(t) \left[ \sigma(t, \bar{x}(t), u) - \sigma(t, \bar{x}(t), \bar{u}(t)) \right) \\
&\quad - \sigma(t, \bar{x}(t), u) - \sigma(t, \bar{x}(t), \bar{u}(t)) \right) \geq 0,
\end{align*}$$

$$\begin{align*}
&\forall u \in U \; \text{a.e.} \; t \in [0, T], \quad \text{a.s.} \; \omega \in \Omega,
\end{align*}$$

with

$$\begin{align*}
&H(t, x, \eta, \zeta) = \langle \eta, b(t, x, u) \rangle + \langle \zeta, \sigma(t, x, u) \rangle, \\
&(t, x, u) \in [0, T] \times \mathbb{R}^n \times U, \quad \eta, \zeta \in \mathbb{R}^n.
\end{align*}$$
This is the well-known maximum principle for the classical stochastic optimal control problem [24], [40]. Note that in the current case, (3.6) is automatic, and (3.5) becomes

\[ P_1(0) \leq 0. \]  

This is a necessary condition in addition to the standard result presented in [40].

2. Optimal control of BSDEs.

Consider the following controlled BSDE:

\[
\begin{aligned}
\frac{dy(t)}{dt} &= g(t, y(t), z(t), u(t)) + z(t)dW(t), \quad t \in [0, T], \\
y(T) &= h_T,
\end{aligned}
\]  

with cost functional

\[ J(u(\cdot)) = f(y(0)). \]  

In this case, we have

\[
\begin{aligned}
b(t, x, y, z, u) &= \sigma(t, x, y, z, u) = 0, \\
g(t, x, y, z, u) &= g(t, y, z, u), \\
\gamma(x_T, y_0) &= 0, \\
h(x_T, y_0) &= h_T, \\
f(x_T, y_0) &= f(y_0).
\end{aligned}
\]

Then, it is easy to see that

\[
\begin{aligned}
x(\cdot) &= 0, \\
\eta(\cdot) &= \zeta(\cdot) = 0, \\
\Phi(\cdot) &= \Psi(\cdot) = 0, \\
P_1(\cdot) &= Q_1(\cdot) = 0, \\
P_2(\cdot) &= Q_2(\cdot) = 0.
\end{aligned}
\]

Consequently, the first adjoint equation (3.1) becomes

\[
\begin{aligned}
\frac{d\xi(t)}{dt} &= -\tilde{g}_y(t)^\top \xi(t) + \tilde{g}_z(t)^\top \xi(t) dt - \tilde{g}_y(t)^\top \xi(t) dW(t), \quad t \in [0, T], \\
\xi(0) &= E\tilde{f}^\top y_0,
\end{aligned}
\]

and the optimality variational inequality (3.7) becomes

\[
\begin{aligned}
\langle \xi(t), g(t, \bar{y}(t), \bar{z}(t), u) - g(t, \bar{y}(t), \bar{z}(t), \bar{u}(t)) \rangle &\geq 0 \\
\forall u \in U, \quad &\text{a.e. } t \in [0, T], \quad &\text{a.s. } \omega \in \Omega.
\end{aligned}
\]

This is essentially the result of [9]. Now, according to our result, we further have the second adjoint equation,

\[
\begin{aligned}
\frac{dP_3(t)}{dt} &= -\left[ \tilde{g}_y(t)^\top P_3(t) + P_3(t)\tilde{g}_y(t) + \sum_{j=1}^m \xi^j(t) g^j_{yy}(t) \right] dt \\
&\quad + Q_3(t) dW(t), \quad t \in [0, T], \\
P_3(T) &= 0,
\end{aligned}
\]

such that the following conditions hold:

\[
\begin{aligned}
E_{\bar{y}_0\bar{y}_0}(\tilde{g}(0)) - P_3(0) &\geq 0, \\
\sum_{j=1}^m \xi^j(t) g^j_{yy}(t) - P_3(t) &\geq 0 \quad &\text{a.e. } t \in [0, T], \quad &\text{a.s.}
\end{aligned}
\]
These are additional necessary conditions, besides those presented in [9].

3. An optimal control for a classical FBSDE.

Consider the following controlled FBSDE:

\[
\begin{aligned}
&dx(t) = b(t, x(t), y(t), z(t), u(t))dt + \sigma(t, x(t), y(t), z(t), u(t))dW(t), \\
&dy(t) = g(t, x(t), y(t), z(t), u(t))dt + z(t)dW(t), \\
&x(0) = \gamma_0, \quad y(T) = h(x(T)),
\end{aligned}
\]

with the cost functional given by the following:

\[
J(u(\cdot)) = \mathbb{E}\left[\varphi(x(T)) + \psi(y(0))\right].
\]

In this case, we have

\[
\gamma(x_T, y_0) = \gamma_0, \quad h(x_T, y_0) = h(x_T), \quad f(x_T, y_0) = \varphi(x_T) + \psi(y_0).
\]

Therefore, we have

\[
\Phi(\cdot) = \Psi(\cdot) = 0.
\]

The first and second adjoint equations become

\[
\begin{aligned}
d\xi(t) &= -\left[\bar{g}_y(t)\xi(t) + \bar{b}_y(t)\eta(t) + \bar{\sigma}_y(t)\zeta(t)\right]dt \\
&\quad - \left[\bar{g}_z(t)\xi(t) + \bar{b}_z(t)\eta(t) + \bar{\sigma}_z(t)\zeta(t)\right]dW(t), \\
d\eta(t) &= -\left[\bar{g}_x(t)\xi(t) + \bar{b}_x(t)\eta(t) + \bar{\sigma}_x(t)\zeta(t)\right]dt + \zeta(t)dW(t), \\
\xi(0) &= \bar{\psi}_{y_0}, \quad \eta(T) = -\bar{h}_{x_T}\xi(T) - \bar{\varphi}_{x_T}
\end{aligned}
\]

and

\[
\begin{aligned}
dP(t) &= -\left[\bar{B}_X(t)^TP(t) + P(t)\bar{B}_X(t) + \bar{\Sigma}_X(t)^TP(t)\bar{\Sigma}_X(t)\right] \\
&\quad + \bar{\Sigma}_X(t)^TQ(t) + Q(t)\bar{\Sigma}_X(t) + \bar{H}_{XX}(t) dt + Q(t)dW(t), \\
P(T) &= -\left[\begin{array}{cc}
\sum_{i=1}^n \xi_i(T)^t\bar{h}_{x_fx_f}^t + \bar{\varphi}_{x_fx_f} & 0 \\
0 & 0\end{array}\right].
\end{aligned}
\]

Condition (3.5) becomes

\[
\begin{pmatrix}
0 & 0 \\
0 & \mathbb{E}\bar{\psi}_{y_0y_0}\end{pmatrix} - P(0) \succeq 0,
\]

and condition (3.6) as well as the optimality variational inequality (3.7) remain the same. When $U$ is convex, we essentially recover the result in [32], and when $\sigma(\cdot)$ is independent of $u(\cdot)$ and $U$ is not assumed to be convex, our result is essentially reduced to that of [30]. Note that in both of these cases, our second adjoint equation is new and leads to some additional necessary conditions.
4. A proof of the main theorem. This section is devoted to a proof of our main theorem. The proof is lengthy and technical. Therefore, we divide it into several steps to make the idea clear.

Step 1. Application of Ekeland’s variation principle and penalized optimal control problem. Let \((x_0, \bar{y}_0, \bar{z}(\cdot), \bar{u}(\cdot))\) be an optimal 4-tuple of Problem \((\hat{C})\), with the corresponding optimal state process \((\hat{x}(\cdot), \hat{y}(\cdot))\). Without loss of generality, we assume that

\[
J(\hat{x}_0, \bar{y}_0, \bar{z}(\cdot), \bar{u}(\cdot)) = 0.
\]

For any \(\delta > 0\), we define

\[
J^\delta(x_0, y_0, z(\cdot), u(\cdot)) = \left\{ \left( J(x_0, y_0, z(\cdot), u(\cdot)) + \delta \right)^+ \right\}^2
\]

\[
+ \mathbb{E} \left[ |x_0 - \gamma(x(T), y_0)|^2 + |y(T) - g(x(T), y_0)|^2 \right] \frac{1}{\delta}
\]

\[
\forall (x_0, y_0, z(\cdot), u(\cdot)) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{Z}^2[0, T] \times \mathcal{U}[0, T].
\]

If we define the usual Ekeland’s distance on \(\mathcal{U}[0, T]\) by

\[
d_E(u(\cdot), \bar{u}(\cdot)) = \int_0^T |\{ t \in [0, T] | u(t) \neq \bar{u}(t) \}| d\mathbb{P} \quad \forall u(\cdot), \bar{u}(\cdot) \in \mathcal{U}[0, T],
\]

with \(|A|\) being the Lebesgue measure of set \(A \subseteq [0, T]\), then \(\mathcal{Z}^2[0, T] \times \mathcal{U}[0, T]\) is a complete metric space under the following metric \([40]\):

\[
\left[ \|z(\cdot) - \bar{z}(\cdot)\|^2 + d_E(u(\cdot), \bar{u}(\cdot))^2 \right]^{\frac{1}{2}} \quad \forall (z(\cdot), u(\cdot)), (\bar{z}(\cdot), \bar{u}(\cdot)) \in \mathcal{Z}^2[0, T] \times \mathcal{U}[0, T],
\]

where \(\|z(\cdot)\|^2 = \mathbb{E} \int_0^T |z(t)|^2 dt\), and \((x_0, y_0, z(\cdot), u(\cdot)) \mapsto J^\delta(x_0, y_0, z(\cdot), u(\cdot))\) is continuous. Also, it is clear that

\[
\begin{cases}
J^\delta(x_0, y_0, z(\cdot), u(\cdot)) > 0, \\
\forall (x_0, y_0, z(\cdot), u(\cdot)) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{Z}^2[0, T] \times \mathcal{U}[0, T], \\
J^\delta(\hat{x}_0, \bar{y}_0, \bar{z}(\cdot), \bar{u}(\cdot)) = \delta \\
\leq \inf_{(x_0, y_0, z(\cdot), u(\cdot)) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{Z}^2[0, T] \times \mathcal{U}[0, T]} J^\delta(x_0, y_0, z(\cdot), u(\cdot)) + \delta.
\end{cases}
\]

Hence, by Ekeland’s variational principle \([20], [40]\), there exists a 4-tuple \((x_0^\delta, y_0^\delta, z^\delta(\cdot), u^\delta(\cdot))\) in \(\mathbb{R}^n \times \mathbb{R}^m \times \mathcal{Z}^2[0, T] \times \mathcal{U}[0, T]\) such that

\[
\begin{cases}
J^\delta(x_0^\delta, y_0^\delta, z^\delta(\cdot), u^\delta(\cdot)) \leq J^\delta(\bar{x}_0, \bar{y}_0, \bar{z}(\cdot), \bar{u}(\cdot)) = \delta, \\
|x_0^\delta - \bar{x}_0|^2 + |y_0^\delta - \bar{y}_0|^2 + \|z^\delta(\cdot) - \bar{z}(\cdot)\|^2 + d_E(u^\delta(\cdot), \bar{u}(\cdot))^2 \leq \delta, \\
-\sqrt{\delta} \left[ |x_0 - x_0^\delta|^2 + |y_0 - y_0^\delta|^2 + \|z(\cdot) - z^\delta(\cdot)\|^2 + d_E(u(\cdot), u^\delta(\cdot))^2 \right]^{\frac{1}{2}} \\
\leq J^\delta(x_0, y_0, z(\cdot), u(\cdot)) - J^\delta(x_0^\delta, y_0^\delta, z^\delta(\cdot), u^\delta(\cdot)), \\
\forall (x_0, y_0, z(\cdot), u(\cdot)) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{Z}^2[0, T] \times \mathcal{U}[0, T].
\end{cases}
\]

Thus, \((x_0^\delta, y_0^\delta, z^\delta(\cdot), u^\delta(\cdot))\) is a global minimum point of the functional

\[
J^\delta(x_0, y_0, z(\cdot), u(\cdot)) + \sqrt{\delta} \left[ |x_0 - x_0^\delta|^2 + |y_0 - y_0^\delta|^2 \\
+ \|z(\cdot) - z^\delta(\cdot)\|^2 + d_E(u(\cdot), u^\delta(\cdot))^2 \right]^{\frac{1}{2}}.
\]
In other words, if we pose an optimal control problem with the state equation (2.6) and the cost functional (4.5), then \((x_0^\delta, y_0^\delta, z^\delta(\cdot), u^\delta(\cdot))\) is an optimal 4-tuple of the problem. We may refer to this problem as a penalized optimal control problem. Note that this problem does not have state constraints, and the optimal 4-tuple \((x_0^\delta, y_0^\delta, z^\delta(\cdot), u^\delta(\cdot))\) approaches \((\bar{x}_0, \bar{y}_0, \bar{z}(\cdot), \bar{u}(\cdot))\) as \(\delta \to 0\).

Step 2. Nontriviality of the multiplier. Note that the state process \((x^\delta(\cdot), y^\delta(\cdot))\) corresponding to the 4-tuple \((x_0^\delta, y_0^\delta, z^\delta(\cdot), u^\delta(\cdot))\) satisfies the following SDE:

\[
\begin{align*}
&dx^\delta(t) = b(t, x^\delta(t), y^\delta(t), z^\delta(t), u^\delta(t))dt \\
&\quad + \sigma(t, x^\delta(t), y^\delta(t), z^\delta(t), u^\delta(t))dW(t), \\
&dy^\delta(t) = g(t, x^\delta(t), y^\delta(t), z^\delta(t), u^\delta(t))dt + z^\delta(t)dW(t), \\
&x^\delta(0) = x_0^\delta, \quad y^\delta(0) = y_0^\delta,
\end{align*}
\]

with

\[
|\gamma^\delta - \gamma(x^\delta(T), y_0^\delta)|^2 + \mathbb{E}[|y^\delta(T) - h(x^\delta(T), y_0^\delta)|^2] \leq J^\delta(x_0^\delta, y_0^\delta, z^\delta(\cdot), u^\delta(\cdot))^2 \leq J^\delta(\bar{x}_0, \bar{y}_0, \bar{z}(\cdot), \bar{u}(\cdot))^2 = \delta^2 \to 0 \quad \text{as} \quad \delta \to 0.
\]

We now regard \(\Theta^\delta(\cdot) \equiv (x^\delta(\cdot), y^\delta(\cdot), z^\delta(\cdot))\) as the unique adapted solution to the following FBSDE:

\[
\begin{align*}
&dx^\delta(t) = b(t, x^\delta(t), y^\delta(t), z^\delta(t), u^\delta(t))dt \\
&\quad + \sigma(t, x^\delta(t), y^\delta(t), z^\delta(t), u^\delta(t))dW(t), \\
&dy^\delta(t) = g(t, x^\delta(t), y^\delta(t), z^\delta(t), u^\delta(t))dt + z^\delta(t)dW(t), \\
&x^\delta(0) = \gamma(x^\delta(0), y^\delta(0)) + \gamma_0^\delta, \quad y^\delta(T) = h(x^\delta(T), y^\delta(0)) + \eta_\delta^T,
\end{align*}
\]

with

\[
\gamma_0^\delta = \left[\begin{array}{c} x_0^\delta - \gamma(x^\delta(T), y_0^\delta) \\ z_0^\delta \end{array}\right], \quad \eta_\delta^T = \left[\begin{array}{c} y^\delta(T) - h(x^\delta(T), y_0^\delta) \end{array}\right].
\]

Next, for any \((\xi_0, \eta_T) \in \mathbb{R}^n \times X_m^2\), with

\[|\xi_0|^2 + \mathbb{E}|\eta_T|^2 \leq 1,\]

by (H0), the following FBSDE admits a unique adapted solution \(\Theta^{\delta, \varepsilon}(\cdot) \equiv (x^{\delta, \varepsilon}(\cdot), y^{\delta, \varepsilon}(\cdot), z^{\delta, \varepsilon}(\cdot))\):

\[
\begin{align*}
&dx^{\delta, \varepsilon}(t) = b(t, x^{\delta, \varepsilon}(t), y^{\delta, \varepsilon}(t), z^{\delta, \varepsilon}(t), u^\delta(t))dt \\
&\quad + \sigma(t, x^{\delta, \varepsilon}(t), y^{\delta, \varepsilon}(t), z^{\delta, \varepsilon}(t), u^\delta(t))dW(t), \\
&dy^{\delta, \varepsilon}(t) = g(t, x^{\delta, \varepsilon}(t), y^{\delta, \varepsilon}(t), z^{\delta, \varepsilon}(t), u^\delta(t))dt + z^{\delta, \varepsilon}(t)dW(t), \\
&x^{\delta, \varepsilon}(0) = \gamma(x^{\delta, \varepsilon}(0), y^{\delta, \varepsilon}(0)) + \gamma_0^\delta + \varepsilon \xi_0 \equiv x_0^{\delta, \varepsilon}, \\
&y^{\delta, \varepsilon}(T) = h(x^{\delta, \varepsilon}(T), y^{\delta, \varepsilon}(0)) + \eta_\delta^T + \varepsilon \eta_T.
\end{align*}
\]

Note that FBSDE (4.10) is nothing but FBSDE (4.8) with only \((\gamma_0^\delta, \eta_\delta^T)\) replaced by \((\gamma_0^\delta + \varepsilon \xi_0, \eta_\delta^T + \varepsilon \eta_T)\), and \(u^\delta(\cdot)\) remains unchanged. Thus, by (2.4), we have

\[
||\Theta^{\delta, \varepsilon}(\cdot) - \Theta^\delta(\cdot)||_{\mathcal{M}^2[0,T]} \leq K\varepsilon^2 \to 0 \quad \text{as} \quad \varepsilon \to 0,
\]
which implies

\[
\lim_{\varepsilon \to 0} \left( |x_{\delta,\varepsilon}^\varepsilon(0) - x_\delta(0)|^2 + |y_{\delta,\varepsilon}^\varepsilon(0) - y_\delta(0)|^2 + \mathbb{E} \int_0^T |z_{\delta,\varepsilon}^\varepsilon(t) - z_\delta(t)|^2 \, dt \right) = 0.
\]

Hence, with \( y_{\delta,\varepsilon}^\varepsilon(0) = y_\delta^\varepsilon(0) \), by taking \((x_0, y_0, z(\cdot), u(\cdot)) = (x_{\delta,\varepsilon}^\varepsilon, y_0^\varepsilon, z_{\delta,\varepsilon}^\varepsilon(\cdot), u_\varepsilon(\cdot)) \) in the last relation in (4.4), we have

\[
-K \sqrt{\delta} \varepsilon \leq J_\delta(x_{\delta,\varepsilon}^\varepsilon, y_{\delta,\varepsilon}^\varepsilon, z_{\delta,\varepsilon}^\varepsilon(\cdot), u_{\varepsilon}(\cdot)) - J_\delta(x_0^\delta, y_0^\delta, z_\delta(\cdot), u_\delta(\cdot))
\]

\[
= \frac{J_\delta(x_{\delta,\varepsilon}^\varepsilon, y_{\delta,\varepsilon}^\varepsilon, z_{\delta,\varepsilon}^\varepsilon(\cdot), u_{\varepsilon}(\cdot))^2 - J_\delta(x_0^\delta, y_0^\delta, z_\delta(\cdot), u_\delta(\cdot))^2}{J_\delta(x_0^\delta, y_0^\delta, z_\delta(\cdot), u_\delta(\cdot))}
\]

\[
= \frac{(J(x_{\delta,\varepsilon}^\varepsilon, y_0^\delta, z_{\delta,\varepsilon}^\varepsilon(\cdot), u_{\delta}(\cdot)) + \delta)^+ - (J(x_0^\delta, y_0^\delta, z_\delta(\cdot), u_\delta(\cdot)) + \delta)^+}{J_\delta(x_0^\delta, y_0^\delta, z_\delta(\cdot), u_\delta(\cdot))}
\]

\[
\geq \Phi_0^{\delta,\varepsilon} \left[ J(x_{\delta,\varepsilon}^\varepsilon, y_0^\delta, z_{\delta,\varepsilon}^\varepsilon(\cdot), u_{\delta}(\cdot)) - J(x_0^\delta, y_0^\delta, z_\delta(\cdot), u_\delta(\cdot)) \right]
\]

\[
+ \langle \Phi_0^{\delta,\varepsilon}, x_{\delta,\varepsilon}^\varepsilon - \gamma(x_{\delta,\varepsilon}^\varepsilon(T), y_{\delta,\varepsilon}^\varepsilon) \rangle
\]

\[
+ \mathbb{E} \langle \Phi_T^{\delta,\varepsilon}(T) - h(x_{\delta,\varepsilon}^\varepsilon(T), y_{\delta,\varepsilon}^\varepsilon) \rangle
\]

\[
\equiv (\Phi_0^{\delta,\varepsilon} + o(1)) \left[ J(x_{\delta,\varepsilon}^\varepsilon, y_0^\delta, z_{\delta,\varepsilon}^\varepsilon(\cdot), u_{\delta}(\cdot)) - J(x_0^\delta, y_0^\delta, z_\delta(\cdot), u_\delta(\cdot)) \right]
\]

\[\epsilon\] stands for certain scalars or vectors that go to 0 as \( \varepsilon \to 0 \),

\[
\Phi_0^{\delta,\varepsilon} = \frac{2 \int_0^1 \beta J(x_{\delta,\varepsilon}^\varepsilon, y_0^\delta, z_{\delta,\varepsilon}^\varepsilon(\cdot), u_\delta(\cdot)) + (1 - \beta) J(x_0^\delta, y_0^\delta, z_\delta(\cdot), u_\delta(\cdot)) + \delta}{J_\delta(x_{\delta,\varepsilon}^\varepsilon, y_0^\delta, z_{\delta,\varepsilon}^\varepsilon(\cdot), u_\delta(\cdot))}
\]

\[
\Phi_T^{\delta,\varepsilon} = \frac{y_{\delta,\varepsilon}^\varepsilon(T) - h(x_{\delta,\varepsilon}^\varepsilon(T), y_{\delta,\varepsilon}^\varepsilon)}{J_\delta(x_{\delta,\varepsilon}^\varepsilon, y_0^\delta, z_{\delta,\varepsilon}^\varepsilon(\cdot), u_\delta(\cdot))}
\]

\[
\Phi_T^{\delta,\varepsilon} \in L^2_T(\Omega; \mathbb{R}^m) \equiv \mathcal{X}_m^2
\]

\[
\Phi_0^{\delta,\varepsilon} \in [0, 1],
\]

where \( o(1) \) stands for certain scalars or vectors that go to 0 as \( \varepsilon \to 0 \).
We point out that \((\Phi_0^\delta, \Phi_0^\epsilon, \Phi^\delta)\) is independent of \((\xi_0, \eta_T)\), and
\[
(4.16) \quad \Phi_0^\delta \geq 0, \quad |\Phi_0^\delta|^2 + |\Phi_0^\epsilon|^2 + \mathbb{E}|\Phi^\delta|^2 = 1.
\]
Thus, we may assume that along some sequence \(\delta \to 0\),
\[
(4.17) \quad \lim_{\delta \to 0} (\Phi_0^\delta, \Phi_0^\epsilon, \Phi_T^\delta) = (\Phi_0, \Phi_0, \Phi_T) \quad \text{weakly in } \mathbb{R} \times \mathbb{R}^n \times \mathcal{A}_m^2.
\]
We claim that
\[
(4.18) \quad \Phi_0 \neq 0.
\]
To show this, let us observe that by (4.11), one has
\[
(4.19) \quad |\Phi_0^\delta| \leq \mathbb{E}(\Phi_T^\delta, \eta_T),
\]
with the constant \(K\) independent of \((\xi_0, \eta_T)\). Now, if \(\Phi_0 = 0\), then the above leads to
\[
(4.20) \quad -r^\delta \leq (\tilde{\Phi}_0^\delta, \xi_0) + \mathbb{E}(\tilde{\Phi}_T^\delta, \eta_T)
\]
with \(r^\delta \to 0\) as \(\delta \to 0\), uniformly in \(|\xi_0|^2 + \mathbb{E}|\eta_T|^2 \leq 1\). Then
\[
(4.21) \quad \left( |\Phi_0^\delta|^2 + \mathbb{E}|\Phi^\delta|^2 \right)^{\frac{1}{2}} = \sup_{|\xi_0|^2 + \mathbb{E}|\eta_T|^2 \leq 1} \left[ (\tilde{\Phi}_0^\delta, \xi_0) + \mathbb{E}(\tilde{\Phi}_T^\delta, \eta_T) \right]
\]
Hence, for \(\delta > 0\) small enough, we must have
\[
1 = |\Phi_0^\delta|^2 + |\Phi_0^\epsilon|^2 + \mathbb{E}|\Phi_T^\delta|^2 < 1,
\]
a contradiction, proving (4.18).

We refer to \((\Phi_0, \Phi_0, \Phi_T)\) as the \textit{Lagrange multiplier} of the corresponding optimal 4-tuple \((\bar{x}_0, \bar{y}_0, \bar{z}(), \bar{u}())\). The nontriviality of \((\Phi_0, \Phi_0, \Phi_T)\) is one of the most important parts in the optimality variational principle, without which the principle is meaningless or useless. We have seen that in the above proof, (H0) plays a crucial role, in a very natural way.

\textit{Step 3. Variations.} For notational simplicity, we now denote
\[
\begin{align*}
X &= \left( \begin{array}{c} x \\ y \\ z \\ u \end{array} \right), & v &= \left( \begin{array}{c} z \\ u \end{array} \right), & X_0 &= \left( \begin{array}{c} x_0 \\ y_0 \end{array} \right), & X_T &= \left( \begin{array}{c} x_T \\ y_T \end{array} \right), \\
B(t, X, v) &= \left( \begin{array}{c} b(t, x, y, z, u) \\ g(t, x, y, z, u) \end{array} \right), & \Sigma(t, X, v) &= \left( \begin{array}{c} \sigma(t, x, y, z, u) \\ z \end{array} \right), \\
F(X_0, X_T) &= f(x_T, y_0), & \Gamma(X_0, X_T) &= \left( \begin{array}{c} x_0 - \gamma(x_T, y_0) \\ y_T - h(x_T, y_0) \end{array} \right), \\
\mathbb{R}^\ell &= \mathbb{R}^\ell \times \mathbb{R}^m, & \mathbb{V}[0, T] &= \mathcal{L}^2[0, T] \times \mathcal{U}[0, T], \\
\mathcal{H} &= \mathbb{R}^\ell \times L^2_{\mathcal{F}_T}(\Omega; \mathbb{R}^\ell) \equiv \mathbb{R}^\ell \times \mathcal{A}_m^2, \\
\mathcal{H}_0 &= \mathbb{R}^\ell \times L^2_{\mathcal{F}_T}(\Omega; \mathbb{R}^m) \equiv \mathbb{R}^\ell \times \mathcal{A}_m^2.
\end{align*}
\]
Consequently,

\begin{align}
J(x_0, y_0, z(\cdot), u(\cdot)) &= J(X_0, v(\cdot)), \\
J^\delta(x_0, y_0, z(\cdot), u(\cdot)) &= J^\delta(X_0, v(\cdot)).
\end{align}

Note that \( \mathcal{H} \) and \( \mathcal{H}_0 \) are Hilbert spaces. We identify \( \mathcal{H}^* = \mathcal{H} \) and \( \mathcal{H}_0^* = \mathcal{H}_0 \). Also, from the above,

\[ F : \mathcal{H} \to \mathbb{R}, \quad \Gamma : \mathcal{H} \to \mathcal{H}_0. \]

We denote the gradient \( DF \) and the Hessian \( D^2F \) of \( F \) as follows:

\[ DF(X_0, X_T) = (F_{X_0}(X_0, X_T), F_{X_T}(X_0, X_T)) \in \mathcal{L}(\mathcal{H}; \mathbb{R}) \equiv \mathcal{H}^* = \mathcal{H}, \]

\[ D^2F(X_0, X_T) = \begin{pmatrix}
F_{X_0X_0}(X_0, X_T) & F_{X_0X_T}(X_0, X_T) \\
F_{X_TX_0}(X_0, X_T) & F_{X_TX_T}(X_0, X_T)
\end{pmatrix} \in \mathcal{L}_s(\mathcal{H}; \mathcal{H}), \]

where \( \mathcal{L}(\mathcal{H}_1; \mathcal{H}_2) \) is the set of all linear bounded operators from \( \mathcal{H}_1 \) to \( \mathcal{H}_2 \), and \( \mathcal{L}_s(\mathcal{H}; \mathcal{H}) \) is the set of all linear bounded self-adjoint operators from \( \mathcal{H} \) to itself. Clearly,

\[
\begin{cases}
F_{X_0}(X_0, X_T) = (0, f_{y_0}) \in \mathbb{R}^t, \\
F_{X_T}(X_0, X_T) = (f_{x_T}, 0) \in \mathcal{X}_T^2, \\
F_{X_0X_0}(X_0, X_T) = \begin{pmatrix} 0 & 0 \\ 0 & f_{y_0} \end{pmatrix} \in \mathcal{S}^{m+n}, \\
F_{X_0X_T}(X_0, X_T) = \begin{pmatrix} f_{x_Ty_0} & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{L}(\mathcal{X}_T^2; \mathcal{X}_T^2), \\
F_{X_TX_0}(X_0, X_T) = \begin{pmatrix} 0 & 0 \\ f_{y_0x_T} & 0 \end{pmatrix} \in \mathcal{L}(\mathcal{X}_T^2; \mathbb{R}^t), \\
F_{X_TX_T}(X_0, X_T) = \begin{pmatrix} 0 & f_{x_Ty_0} \\ 0 & 0 \end{pmatrix} \in \mathcal{L}(\mathbb{R}^t, \mathcal{X}_T^2). 
\end{cases}
\]

Now, we look at \( \Gamma : \mathcal{H} \to \mathcal{H}_0 \). By the Fréchet differentiability, we have

\[ D\Gamma(X_0, X_T) = (\Gamma_{X_0}(X_0, X_T), \Gamma_{X_T}(X_0, X_T)) \in \mathcal{L}(\mathcal{H}; \mathcal{H}_0) \]

and

\[ D^2\Gamma(X_0, X_T) = \begin{pmatrix}
\Gamma_{X_0X_0}(X_0, X_T) & \Gamma_{X_0X_T}(X_0, X_T) \\
\Gamma_{X_TX_0}(X_0, X_T) & \Gamma_{X_TX_T}(X_0, X_T)
\end{pmatrix} \in \mathcal{L}(\mathcal{H}; \mathcal{L}(\mathcal{H}; \mathcal{H}_0)). \]

To make the above more precise, let us take any \( \Phi \equiv (\hat{\Phi}_0, \hat{\Phi}_T) \in \mathcal{H}_0 \). Then

\[
\langle \Gamma(X_0, X_T), \Phi \rangle = \langle x_0 - \gamma(x_T, y_0), \hat{\Phi}_0 \rangle + \mathbb{E} \langle y_T - h(x_T, y_0), \hat{\Phi}_T \rangle \\
= \sum_{i=1}^n \hat{\Phi}_0^i [x_0^i - \gamma_i(x_T, y_0)] + \mathbb{E} \left( \sum_{j=1}^m \hat{\Phi}_T^j [y_T^j - h_i(x_T, y_0)] \right).
\]

Thus,

\[
[D\Gamma(X_0, X_T)\hat{\Phi}] = D[\langle \Gamma(X_0, X_T), \hat{\Phi} \rangle] \\
= \left( \langle \Gamma(X_0, X_T), \hat{\Phi} \rangle \right)_{X_0} \langle \Gamma(X_0, X_T), \hat{\Phi} \rangle_{X_T} \\
= \left( \Gamma_{X_0}(X_0, X_T)\hat{\Phi}, \Gamma_{X_T}(X_0, X_T)\hat{\Phi} \right) \in \mathcal{L}(\mathcal{H}; \mathbb{R}),
\]
with
\[
\Gamma_{X_0}(X_0, X_T) \Phi = \langle \Gamma(X_0, X_T), \Phi \rangle_{X_0} = \left( \Phi_0^T - \Phi_0^T \gamma_{y_0}(x_T, y_0) - E[\Phi_T^T h_{y_0}(x_T, y_0)] \right),
\]
\[
\Gamma_{X_T}(X_0, X_T) \Phi = \langle \Gamma(X_0, X_T), \Phi \rangle_{X_T} = \left( - \Phi_0^T \gamma_{x_T}(x_T, y_0) - \Phi_T^T h_{x_T}(x_T, y_0), \Phi_T \right),
\]
and
\[
[D^2 \Gamma(X_0, X_T) \Phi] = D^2[\langle \Gamma(X_0, X_T), \Phi \rangle] = \left( \begin{array}{cc}
\langle \Gamma(X_0, X_T), \Phi \rangle_{X_0} & \langle \Gamma(X_0, X_T), \Phi \rangle_{X_0 X_T} \\
\langle \Gamma(X_0, X_T), \Phi \rangle_{X_T X_0} & \langle \Gamma(X_0, X_T), \Phi \rangle_{X_T X_T}
\end{array} \right).
\]

Now, fixed \( X_0 \in \mathbb{R}^\ell \), and \( v(\cdot) \in \mathcal{V}[0, T] \). For any \( \varepsilon \in (0, 1) \), let
\[
X_0^{\delta, \varepsilon} = X_0^{\delta, \varepsilon} + \sqrt{\varepsilon} X_0
\]
and
\[
v^{\delta, \varepsilon}(t) = \begin{cases} v^{\delta}(t), & t \in [0, T] \setminus S_{\varepsilon}, \\ v(t), & t \in S_{\varepsilon} \end{cases}
\]
for some \( S_{\varepsilon} \subseteq [0, T] \), undetermined, with \( |S_{\varepsilon}| = \varepsilon T \) (recall that \( |S_{\varepsilon}| \) is the Lebesgue measure of \( S_{\varepsilon} \)). Let \( X^{\delta, \varepsilon}(\cdot) \) be the state process corresponding to \( (X_0^{\delta, \varepsilon}, v^{\delta, \varepsilon}(\cdot)) \). Further, let \( X_1^{\delta, \varepsilon}(\cdot) \) and \( X_2^{\delta, \varepsilon}(\cdot) \) be the solutions to the following SDEs:
\[
\begin{cases}
\frac{dX_1^{\delta, \varepsilon}(t)}{dt} = B^{\delta}(t)X_1^{\delta, \varepsilon}(t) + \sum_{j=1}^{m} \lambda_j^2(t)X_1^{\delta, \varepsilon}(t) + \Delta \Sigma^2(t) I_{S_{\varepsilon}}(t) dW(t), \\
X_1^{\delta, \varepsilon}(0) = \sqrt{\varepsilon} X_0
\end{cases}
\]
and

\[
\begin{cases}
\ dX_2^{\delta,\varepsilon}(t) = \left[ B_X^\delta(t)X_1^\varepsilon(t) + \left( \Delta B_1^\delta(t) + \Delta B_X^\delta(t)X_1^\delta(t) \right) I_{S_e}(t) \right. \\
\ + \frac{1}{2}B_X^\delta(t)X_1^{\delta,\varepsilon}(t)^2 \right] dt \\
\ + \left[ \Sigma_X^\delta(t)X_2^{\delta,\varepsilon}(t) + \Delta \Sigma_X^\delta(t)X_1^{\delta,\varepsilon}(t)I_{S_e}(t) \right. \\
\ + \frac{1}{2}\Sigma_X^\delta(t)X_1^{\delta,\varepsilon}(t)^2 \right] dW(t), \\
\ X_2^{\delta,\varepsilon}(0) = 0,
\end{cases}
\]

(4.26)

where (recall (4.21))

\[
\begin{align*}
B_X^\delta(t) &= B_X(t, X^\delta(t), v^\delta(t)), & \Sigma_X^\delta(t) &= \Sigma_X(t, X^\delta(t), v^\delta(t)), \\
\Delta B^\delta(t) &= B(t, X^\delta(t), v(t)) - B(t, X^\delta(t), v^\delta(t)), \\
\Delta \Sigma_X^\delta(t) &= \Sigma(t, X^\delta(t), v(t)) - \Sigma(t, X^\delta(t), v^\delta(t)), \\
\Delta B_X^\delta(t) &= B_X(t, X^\delta(t), v(t)) - B_X(t, X^\delta(t), v^\delta(t)), \\
\Delta \Sigma_X^\delta(t) &= \Sigma_X(t, X^\delta(t), v(t)) - \Sigma_X(t, X^\delta(t), v^\delta(t)).
\end{align*}
\]

We have the following estimates, whose proofs are similar to those given in [22]. For any \( k \geq 1 \),

\[
\begin{align*}
&\mathbb{E} \left[ \sup_{t \in [0, T]} |X_{1,\varepsilon}^\delta(t)|^{2k} \right] + \mathbb{E} \left[ \sup_{t \in [0, T]} |X_{1,\varepsilon}^\delta(t) - X^\delta(t)|^{2k} \right] \leq C\varepsilon^k, \\
&\mathbb{E} \left[ \sup_{t \in [0, T]} |X_{1,\varepsilon}^\delta(t) - X_1^{\delta,\varepsilon}(t)|^{2k} \right] + \mathbb{E} \left[ \sup_{t \in [0, T]} |X_2^{\delta,\varepsilon}(t)|^{2k} \right] \leq C\varepsilon^{2k}, \\
&\mathbb{E} \left[ \sup_{t \in [0, T]} |X_{1,\varepsilon}^\delta(t) - X_1^{\delta,\varepsilon}(t) - X_2^{\delta,\varepsilon}(t)|^{2k} \right] = o(\varepsilon^{2k}).
\end{align*}
\]

(4.27)

We point out that everything is reduced to the classical optimal control for SDEs, and that the above estimates are better than those presented in [40]. Now, by (4.4),
we obtain

\[-\sqrt{\delta}\left(\sqrt{\varepsilon}X_0 + \varepsilon T\right) \leq J^\delta(X^\delta_{\delta}, \nu^\delta(\cdot)) - J^\delta(X^\delta_{0}, \nu^\delta(\cdot))\]

\[
= J^\delta(X^\delta_{\delta}, \nu^\delta(\cdot))^2 - J^\delta(X^\delta_{0}, \nu^\delta(\cdot))^2 \\
= J^\delta(X^\delta_{\delta}, \nu^\delta(\cdot)) + J^\delta(X^\delta_{0}, \nu^\delta(\cdot)) \\
\left[ (J(X^\delta_{\delta}, \nu^\delta(\cdot)) + \delta)^\frac{1}{2} - (J(X^\delta_{0}, \nu^\delta(\cdot)) + \delta)^\frac{1}{2}\right]^2 \\
= J^\delta(X^\delta_{\delta}, \nu^\delta(\cdot)) + J^\delta(X^\delta_{0}, \nu^\delta(\cdot)) \\
+ \mathbb{E}\left[ \Gamma(X^\delta_{\delta}, X^\delta_{\delta}(T)) - \Gamma(X^\delta_{0}, X^\delta(T))\right] \\
= \Phi^\delta + \mathbb{E}\left[ \Phi^\delta, \Gamma(X^\delta_{\delta}, X^\delta_{\delta}(T)) - \Gamma(X^\delta_{0}, X^\delta(T))\right] \\
\equiv \Phi^\delta_0 + \mathbb{E}\left[ \Phi^\delta + \mathbb{E}\left( \Phi^\delta, \Gamma(X^\delta_{\delta}, X^\delta_{\delta}(T)) - \Gamma(X^\delta_{0}, X^\delta(T))\right)\right],
\]

where \((\Phi^\delta_0, \bar{\Phi}^\delta_0, \bar{\Phi}^\delta_T)\) is defined similarly to (4.15), and \((\Phi^\delta_0, \bar{\Phi}^\delta_0, \bar{\Phi}^\delta_T)\) is defined similarly to (4.14), with \(X^\delta_{\delta} = \left(\delta^\delta, \delta^\delta\right)\) replaced by (4.23). We have shown that along a sequence,

\[(\Phi^\delta_0, \bar{\Phi}^\delta) \to (\Phi_0, \bar{\Phi}), \quad \Phi^0 \neq 0, \quad \bar{\Phi} = (\bar{\Phi}_0, \bar{\Phi}_T).
\]

Note that

\[
\begin{align*}
J(X^\delta_{\delta}, \nu^\delta(\cdot)) - J(X^\delta_{0}, \nu^\delta(\cdot)) \\
= \mathbb{E}\left[ F(X^\delta_{\delta}, X^\delta_{\delta}(T)) - F(X^\delta_{0}, X^\delta(T))\right] \\
= \mathbb{E}\left\{ D_F^\delta \left(X^\delta_{\delta} - X^\delta\right) + \frac{1}{2} D_F^\delta \left(X^\delta_{\delta} - X^\delta\right) \cdot \left(X^\delta_{\delta} - X^\delta\right) \cdot \left(X^\delta_{\delta} - X^\delta\right) \cdot \left(X^\delta_{\delta} - X^\delta\right) \cdot \left(X^\delta_{\delta} - X^\delta\right) \right\} \\
= \mathbb{E}\left\{ D_F^\delta [X^\delta_{\delta} - X^\delta_{0}] + F_X^\delta [X^\delta_{\delta}(T) - X^\delta(T)] \right\}
\end{align*}
\]
where

\[
\begin{align*}
D^F &= \left( F_{X_T} - F_{X_0} \right) \equiv \left( F_{X_0} (X_0^\delta, X^\delta(T)), F_{X_T} (X_0^\delta, X^\delta(T)) \right), \\
D^2 F &= \left( F_{X_0 X_0} - F_{X_0 X_T} \right) \equiv \left( F_{X_0 X_0} (X_0^\delta, X^\delta(T)), F_{X_0 X_T} (X_0^\delta, X^\delta(T)) \right), \\
D^2 F_{\delta, \varepsilon} &= \int_0^1 \beta \bigg( D^2 F (\beta X_0^\delta + (1 - \beta) X_0^{\delta, \varepsilon}, \beta X^\delta(T) + (1 - \beta) X^{\delta, \varepsilon}(T)) \\
&\quad - D^2 F (X_0^\delta, X^\delta(T)) \bigg) d\beta.
\end{align*}
\]

Also, we have

\[
\mathbb{E} \left\langle \Phi^{\delta, \varepsilon}, \Gamma (X_0^{\delta, \varepsilon}, X^{\delta, \varepsilon}(T)) - \Gamma (X_0^\delta, X^\delta(T)) \right\rangle
= \mathbb{E} \left\langle D \Phi^{\delta, \varepsilon} \cdot \left( X^\delta - X_0^{\delta, \varepsilon} \right), X^{\delta, \varepsilon}(T) - X^\delta(T) \right\rangle
= \mathbb{E} \left\langle D \Gamma (X_0^\delta, X^\delta(T)) \Phi^{\delta, \varepsilon}, \right. \\
&\quad \left. + \frac{1}{2} \left( D^2 \Gamma_{\delta, \varepsilon} \Phi^{\delta, \varepsilon} \cdot \left( X^\delta - X_0^{\delta, \varepsilon} \right), X^{\delta, \varepsilon}(T) - X^\delta(T) \right) \right\rangle
\]

where

\[
\begin{align*}
\left[ D \Gamma \Phi^{\delta, \varepsilon} \right] &= D \Gamma (X_0^\delta, X^\delta(T)) \Phi^{\delta, \varepsilon}, \\
\left[ D^2 \Gamma \Phi^{\delta, \varepsilon} \right] &= D^2 \Gamma (X_0^\delta, X^\delta(T)) \Phi^{\delta, \varepsilon},
\end{align*}
\]

and

\[
\begin{align*}
\left[ \Gamma_{\delta, X_0} \Phi^{\delta, \varepsilon} \right] &= \Gamma_{\delta, X_0} (X_0^\delta, X^\delta(T)) \Phi^{\delta, \varepsilon}, \\
\left[ \Gamma_{\delta, X_T} \Phi^{\delta, \varepsilon} \right] &= \Gamma_{\delta, X_T} (X_0^\delta, X^\delta(T)) \Phi^{\delta, \varepsilon}, \\
\left[ \Gamma_{\delta, X_0 X_0} \Phi^{\delta, \varepsilon} \right] &= \Gamma_{\delta, X_0 X_0} (X_0^\delta, X^\delta(T)) \Phi^{\delta, \varepsilon}, \\
\left[ \Gamma_{\delta, X_0 X_T} \Phi^{\delta, \varepsilon} \right] &= \Gamma_{\delta, X_0 X_T} (X_0^\delta, X^\delta(T)) \Phi^{\delta, \varepsilon}, \\
\left[ \Gamma_{\delta, X_T X_0} \Phi^{\delta, \varepsilon} \right] &= \Gamma_{\delta, X_T X_0} (X_0^\delta, X^\delta(T)) \Phi^{\delta, \varepsilon}, \\
\left[ \Gamma_{\delta, X_T X_T} \Phi^{\delta, \varepsilon} \right] &= \Gamma_{\delta, X_T X_T} (X_0^\delta, X^\delta(T)) \Phi^{\delta, \varepsilon}, \\
\left[ D^2 \Gamma_{\delta, \varepsilon} \Phi^{\delta, \varepsilon} \right] &= \int_0^1 \beta \left[ D^2 \Gamma (\beta X_0^\delta + (1 - \beta) X_0^{\delta, \varepsilon}, \beta X^\delta(T) + (1 - \beta) X^{\delta, \varepsilon}(T)) \Phi^{\delta, \varepsilon} \\
&\quad - D^2 \Gamma (X_0^\delta, X^\delta(T)) \Phi^{\delta, \varepsilon} \right] d\beta.
\end{align*}
\]
Under our conditions, we have
\[ |\Phi_0^{\delta,\varepsilon} D^{\delta,\varepsilon} | + |D^2 \Gamma^{\delta,\varepsilon} \Phi^{\delta,\varepsilon} | \leq K \left( |X_0^{\delta,\varepsilon} - X_0^\delta| + |X^{\delta,\varepsilon}(T) - X^\delta(T)| \right). \]

Hence,
\[
\mathbb{E} \left\{ \left| \Phi_0^{\delta,\varepsilon} D^{\delta,\varepsilon} \left( X_0^{\delta,\varepsilon} - X_0^\delta \right) \cdot \left( X^{\delta,\varepsilon}(T) - X^\delta(T) \right) \right| \\
+ \left| \left[ D^2 \Gamma^{\delta,\varepsilon} \Phi^{\delta,\varepsilon} \right] \left( X_0^{\delta,\varepsilon} - X_0^\delta \right) \cdot \left( X^{\delta,\varepsilon}(T) - X^\delta(T) \right) \right| \right\} \\
\leq K \mathbb{E} \left( |X_0^{\delta,\varepsilon} - X_0^\delta| + |X^{\delta,\varepsilon}(T) - X^\delta(T)| \right) \leq K \varepsilon^{\frac{3}{2}}.
\]

Consequently,
\[
-\sqrt{\delta} \left( \sqrt{\varepsilon} |X_0| + \varepsilon T \right) \\
\leq \Phi_0^{\delta,\varepsilon} \left[ J(X_0^{\delta,\varepsilon}, v^{\delta,\varepsilon}(\cdot)) - J(X_0^\delta, v^\delta(\cdot)) \right] \\
+ \mathbb{E} \left( \langle \Phi^{\delta,\varepsilon}, \Gamma(X_0^{\delta,\varepsilon}, X^{\delta,\varepsilon}(T)) - \Gamma(X_0^\delta, X^\delta(T)) \rangle \right) \\
= \mathbb{E} \left\{ \Phi_0^{\delta,\varepsilon} \left[ F_{X_0}^\delta |X_0^{\delta,\varepsilon} - X_0^\delta| + F_{X_T}^\delta |X^{\delta,\varepsilon}(T) - X^\delta(T)| \right] \\
+ \frac{1}{2} \left( F_{X_0}^\delta X_0^{\delta,\varepsilon} - X_0^\delta, X_0^{\delta,\varepsilon} - X_0^\delta \right) \\
+ \left( F_{X_T}^\delta X_T^{\delta,\varepsilon} - X_T^\delta, X_T^{\delta,\varepsilon} - X_T^\delta \right) \\
+ \frac{1}{2} \left( F_{X_T}^\delta X_T^{\delta,\varepsilon} - X_T^\delta, X_T^{\delta,\varepsilon}(T) - X_T^\delta(T) \right) \right\} \\
+ O(\varepsilon^{\frac{3}{2}}) \\
= \mathbb{E} \left\{ \sqrt{\varepsilon} \left( \Phi_0^{\delta,\varepsilon} F_{X_0}^\delta \right)^\top + \Gamma_{X_0}^\delta \left( \Phi^{\delta,\varepsilon} \right), X_0 \right\} \\
+ \langle \Phi_0^{\delta,\varepsilon} F_{X_0}^\delta, X_1^{\delta,\varepsilon}(T) + X_2^{\delta,\varepsilon}(T) \rangle \\
+ \frac{\varepsilon}{2} \left( \Phi_0^{\delta,\varepsilon} F_{X_0}^\delta + \Gamma_{X_0}^\delta \left( \Phi^{\delta,\varepsilon} \right) \right) X_0, X_0 \right\} \\
+ \sqrt{\varepsilon} \left\{ \Phi_0^{\delta,\varepsilon} F_{X_T}^\delta + \Gamma_{X_T}^\delta \left( \Phi^{\delta,\varepsilon} \right) \right\} X_T^{\delta,\varepsilon}(T), X_T^{\delta,\varepsilon}(T) \right\} + O(\varepsilon^{\frac{3}{2}}).
\]

**Step 4. Duality.** Let \( (\hat{\Phi}^{\delta,\varepsilon}(\cdot), \hat{\Psi}^{\delta,\varepsilon}(\cdot)) \) be the adapted solution to the following BSDE:

\[
\begin{align*}
\langle d\hat{\Phi}^{\delta,\varepsilon}(t) \rangle &= -B_X(t)^\top \hat{\Phi}^{\delta,\varepsilon}(t) + \Sigma_X(t)^\top \hat{\Psi}^{\delta,\varepsilon}(t) \right dt + \hat{\Phi}^{\delta,\varepsilon}(t) dW(t), \\
\hat{\Phi}^{\delta,\varepsilon}(T) &= -\left( \Phi_0^{\delta,\varepsilon}(F_{X_T}^\delta)^\top + \left[ \Gamma_{X_T}^\delta, \hat{\Phi}^{\delta,\varepsilon} \right] \right) X_T^{\delta,\varepsilon}(T), X_T^{\delta,\varepsilon}(T) \right) \right\} + O(\varepsilon^{\frac{3}{2}}).
\end{align*}
\]
and let \((\Phi^{\delta, \epsilon}(\cdot), \Psi^{\delta, \epsilon}(\cdot))\) be the unique adapted solution to the following:

\[
\begin{align*}
(4.32) & \quad \left\{ \begin{array}{l}
d\Phi^{\delta, \epsilon}(t) = -\left[ B_X^\delta(t)^\top \Phi^{\delta, \epsilon}(t) + \Sigma_X^\delta(t)^\top \Psi^{\delta, \epsilon}(t) \right] dt + \Psi^{\delta, \epsilon}(t)dW(t), \\
\Phi^{\delta, \epsilon}(T) = -\left\{ \Phi_0^{\delta, \epsilon} F_{X^\epsilon T} X_0 + \left[ \Gamma_{X^\epsilon T} X_0 \Phi^{\delta, \epsilon} \right] \right\}.
\end{array} \right.
\]

Then as \((\epsilon, \delta) \to (0, 0)\), \((\Phi^{\delta, \epsilon}(\cdot), \Psi^{\delta, \epsilon}(\cdot))\) goes to \((\Phi(\cdot), \Psi(\cdot))\) and \((\Phi^{\delta, \epsilon}(\cdot), \Psi^{\delta, \epsilon}(\cdot))\) goes to \((\Phi(\cdot), \Psi(\cdot))\), which are, respectively, the adapted solutions of the following BSDEs:

\[
(4.33) \quad \left\{ \begin{array}{l}
d\Phi(t) = -\left[ B_X(t)^\top \Phi(t) + \Sigma_X(t)^\top \Psi(t) \right] dt + \Psi(t)dW(t), \\
\Phi(T) = -\left\{ \Phi_0 F_{X T} X_0 + \left[ \Gamma_{X T} \Phi \right] \right\}
\end{array} \right.
\]

and

\[
(4.34) \quad \left\{ \begin{array}{l}
d\Phi(t) = -\left[ B_X(t)^\top \Phi(t) + \Sigma_X(t)^\top \Psi(t) \right] dt + \Psi(t)dW(t), \\
\Phi(T) = -\left\{ \Phi_0 F_{X T} X_0 + \left[ \Gamma_{X T} \Phi \right] \right\}.
\end{array} \right.
\]

Consequently,

\[
-\mathbb{E} \left\{ \Phi_0^{\delta, \epsilon} (F_{X^\epsilon T})^\top + \left[ \Gamma_{X^\epsilon T} \Phi^{\delta, \epsilon} \right], \Phi^{\delta, \epsilon}(T) + X_1^{\delta, \epsilon}(T) + X_2^{\delta, \epsilon}(T) \right\}
= \mathbb{E} \left\{ \Phi^{\delta, \epsilon}(T), X_1^{\delta, \epsilon}(T) + X_2^{\delta, \epsilon}(T) \right\}
= \mathbb{E} \left\{ \Phi^{\delta, \epsilon}(0), \sqrt{\epsilon} X_0 \right\} + \int_0^T \left[ \left( -B_X^\delta(t)^\top \Phi^{\delta, \epsilon}(t) - \Sigma_X^\delta(t)^\top \Psi^{\delta, \epsilon}(t), X_1^{\delta, \epsilon}(t) + X_2^{\delta, \epsilon}(t) \right) \right.
+ \left\{ \Phi^{\delta, \epsilon}(t), B_X^\delta(t)[X_1^{\delta, \epsilon}(t) + X_2^{\delta, \epsilon}(t)] \right.
+ \left[ \Delta B^\delta(t) + \Delta B^\delta_X(t) X_1^{\delta, \epsilon}(t) \right] I_{S_{\epsilon}}(t) + \frac{1}{2} B_X^\delta(t) X_1^{\delta, \epsilon}(t)^2 \right] dt
+ \left\{ \Psi^{\delta, \epsilon}(t), \Sigma_X^\delta(t)[X_1^{\delta, \epsilon}(t) + X_2^{\delta, \epsilon}(t)] + [\Delta \Sigma^\delta(t) + \Delta \Sigma_X^\delta(t) X_1^{\delta, \epsilon}(t)] I_{S_{\epsilon}}(t)
+ \frac{1}{2} \Sigma_X^\delta(t) X_1^{\delta, \epsilon}(t)^2 \right\} \right] dt
= \mathbb{E} \left\{ \Phi^{\delta, \epsilon}(0), \sqrt{\epsilon} X_0 \right\} + \int_0^T \left[ \left( \Phi^{\delta, \epsilon}(t), \Delta B^\delta(t) I_{S_{\epsilon}}(t) + \frac{1}{2} B_X^\delta(t) X_1^{\delta, \epsilon}(t)^2 \right) \right.
+ \left\{ \Psi^{\delta, \epsilon}(t), \Delta \Sigma^\delta(t) I_{S_{\epsilon}}(t) + \frac{1}{2} \Sigma_X^\delta(t) X_1^{\delta, \epsilon}(t)^2 \right\} \right] dt
+ O(\epsilon^2)
= \mathbb{E} \left\{ \Phi^{\delta, \epsilon}(0), \sqrt{\epsilon} X_0 \right\} + \int_0^T \left[ \left( \Phi^{\delta, \epsilon}(t), \Delta B^\delta(t) \right) + \left( \Psi^{\delta, \epsilon}(t), \Delta \Sigma^\delta(t) \right) \right] I_{S_{\epsilon}}(t)
+ \frac{1}{2} \left( H_{XX}^\delta(t) X_1^{\delta, \epsilon}(t), X_1^{\delta, \epsilon}(t) \right) \right] dt
+ O(\epsilon^2),
\]

where

\[
H_{XX}^\delta(t) = H_{XX}(t, X^\delta(t), \psi^\delta(t), \Phi^\delta(t), \Psi^\delta(t)),
\]
with $H_{XX}(t, X, v, \tilde{\Phi}, \tilde{\Psi})$ being the Hessian of
\[ H(t, X, v, \tilde{\Phi}, \tilde{\Psi}) = \langle \tilde{\Phi}, B(t, X, v) \rangle + \langle \tilde{\Psi}, \Sigma(t, X, v) \rangle, \]
with respect to $X$. Likewise,
\[ -\varepsilon \left( \langle -\bar{E} X(t), \sqrt{\bar{E} X(t)} \rangle \right) \leq \mathbb{E}\left\{ \langle \Phi_0^{\delta,\varepsilon}(F_{X_0}^{\delta})^\top + [\Gamma_{X_0}^{\delta} \tilde{\Phi}^{\delta,\varepsilon}], \sqrt{\bar{E} X(t)} \rangle \right\} \]
\[-\varepsilon \left( \langle \Phi_0^{\delta,\varepsilon}(0), \sqrt{\bar{E} X(t)} \rangle \right) + \int_0^T \left( \langle \Phi_0^{\delta,\varepsilon}(t), \Delta B(t) \rangle + \langle \tilde{\Psi}^{\delta,\varepsilon}, \Delta \Sigma(t) \rangle I_{S_t}(t) \right) dt \]
\[ + \int_0^T \left( \langle \Phi_0^{\delta,\varepsilon}(t), \Delta B(t) \rangle + \langle \tilde{\Psi}^{\delta,\varepsilon}, \Delta \Sigma(t) \rangle I_{S_t}(t) \right) dt \]
\[ - \varepsilon \langle \Phi_0^{\delta,\varepsilon}(0), X_0 \rangle + \frac{1}{2} \left( \langle \Phi_0^{\delta,\varepsilon}(F_{X_0}^{\delta})^\top + [\Gamma_{X_0}^{\delta} \tilde{\Phi}^{\delta,\varepsilon}], X_0 \rangle \right) \]
\[ = \mathbb{E}\left\{ \langle \Phi_0^{\delta,\varepsilon}(F_{X_0}^{\delta})^\top + [\Gamma_{X_0}^{\delta} \tilde{\Phi}^{\delta,\varepsilon}], \sqrt{\bar{E} X(t)} \rangle \right\} \]
\[ + \int_0^T \left( \langle \Phi_0^{\delta,\varepsilon}(t), \Delta B(t) \rangle + \langle \tilde{\Psi}^{\delta,\varepsilon}, \Delta \Sigma(t) \rangle I_{S_t}(t) \right) dt \]
\[ + \frac{\varepsilon}{2} \left( \langle \Phi_0^{\delta,\varepsilon}(0), X_0 \rangle + \left( \langle \Phi_0^{\delta,\varepsilon}(F_{X_0}^{\delta}) + [\Gamma_{X_0}^{\delta} \tilde{\Phi}^{\delta,\varepsilon}] \right) X_0, X_0 \rangle \right) \]
\[ + \frac{1}{2} \left( \langle \Phi_0^{\delta,\varepsilon}(t), \Delta B(t) \rangle + \langle \tilde{\Psi}^{\delta,\varepsilon}, \Delta \Sigma(t) \rangle I_{S_t}(t) \right) \]
\[ - \frac{1}{2} \int_0^T \langle H_{XX}^{\delta,\varepsilon}(t)X_1^{\delta,\varepsilon}(t), X_1^{\delta,\varepsilon}(t) \rangle dt \right\} + O(\varepsilon^2). \]

Note that $Y^{\delta,\varepsilon}(\cdot) \equiv X_1^{\delta,\varepsilon}(\cdot)X_1^{\delta,\varepsilon}(\cdot)^\top$ satisfies
\[ dY^{\delta,\varepsilon}(t) = \left[ B_X^{\delta}(t)Y^{\delta,\varepsilon}(t) + Y^{\delta,\varepsilon}(t)B_X^{\delta}(t)^\top + \Sigma_X^{\delta}(t)Y^{\delta,\varepsilon}(t)\Sigma_X^{\delta}(t)^\top \right. \]
\[ + \left( \Delta \Sigma^{\delta}(t)X_1^{\delta,\varepsilon}(t)\Delta \Sigma^{\delta}(t)^\top \right) I_{S_t}(t) dt \]
\[ + \left[ \Sigma_X^{\delta}(t)Y^{\delta,\varepsilon}(t) + Y^{\delta,\varepsilon}(t)\Sigma_X^{\delta}(t)^\top \right] \]
\[ + \left( \Delta \Sigma^{\delta}(t)X_1^{\delta,\varepsilon}(t) + X_1^{\delta,\varepsilon}(t)\Delta \Sigma^{\delta}(t)^\top \right) I_{S_t}(t) \right] \]
\[ + \frac{1}{2} \left( \langle \Phi_0^{\delta,\varepsilon}(t), \Delta B(t) \rangle + \langle \tilde{\Psi}^{\delta,\varepsilon}, \Delta \Sigma(t) \rangle I_{S_t}(t) \right) \]
\[ = \mathbb{E}\left\{ \langle \Phi_0^{\delta,\varepsilon}(F_{X_0}^{\delta})^\top + [\Gamma_{X_0}^{\delta} \tilde{\Phi}^{\delta,\varepsilon}], \sqrt{\bar{E} X(t)} \rangle \right\} \]
\[ + \int_0^T \left( \langle \Phi_0^{\delta,\varepsilon}(t), \Delta B(t) \rangle + \langle \tilde{\Psi}^{\delta,\varepsilon}, \Delta \Sigma(t) \rangle I_{S_t}(t) \right) dt \]
\[ + \frac{\varepsilon}{2} \left( \langle \Phi_0^{\delta,\varepsilon}(0), X_0 \rangle + \left( \langle \Phi_0^{\delta,\varepsilon}(F_{X_0}^{\delta}) + [\Gamma_{X_0}^{\delta} \tilde{\Phi}^{\delta,\varepsilon}] \right) X_0, X_0 \rangle \right) \]
\[ + \frac{1}{2} \left( \langle \Phi_0^{\delta,\varepsilon}(t), \Delta B(t) \rangle + \langle \tilde{\Psi}^{\delta,\varepsilon}, \Delta \Sigma(t) \rangle I_{S_t}(t) \right) \]
\[ - \frac{1}{2} \int_0^T \langle H_{XX}^{\delta,\varepsilon}(t)X_1^{\delta,\varepsilon}(t), X_1^{\delta,\varepsilon}(t) \rangle dt \right\} + O(\varepsilon^2). \]
Now, let \((P^{b,\varepsilon}(\cdot), Q^{b,\varepsilon}(\cdot))\) be the adapted solution to the following BSDE:

\[
\begin{align*}
\frac{dP^{b,\varepsilon}(t)}{dt} &= -\left[B^{b,\varepsilon}(t)^\top P^{b,\varepsilon}(t) + P^{b,\varepsilon}(t)B^{b,\varepsilon}(t) + \Sigma^{b,\varepsilon}(t)^\top P^{b,\varepsilon}(t)\Sigma^{b,\varepsilon}(t)\right] dt \\
&\quad + \Sigma^{b,\varepsilon}(t)^\top Q^{b,\varepsilon}(t) + Q^{b,\varepsilon}(t)\Sigma^{b,\varepsilon}(t) + H^{b,\varepsilon}(X(t)) dt \\
&\quad + Q^{b,\varepsilon}(t)dW(t), \\
P^{b,\varepsilon}(T) &= -\left\{\Phi_{0,\varepsilon}^{b} F_{X_T X_T} + [\Gamma_{X_T X_T} \Phi_{0,\varepsilon}^{b}]\right\}. 
\end{align*}
\]

(4.36)

Then, by \([40]\), we have

\[
\begin{align*}
\epsilon \to 0, \quad &\Rightarrow \\
\Rightarrow &\text{ (4.40)} \\
\end{align*}
\]

\[
\begin{align*}
\text{Next, dividing (4.39), and then sending } |X_0| \to \infty, \quad &\text{and then sending } |X_0| \to \infty, \quad \epsilon \to 0, \quad \delta \to 0, \text{ we see that} \\
&\text{we see that}
\end{align*}
\]

\[
\begin{align*}
\begin{cases}
&\Phi(0) \leq \mathbb{E}\left\{\Phi_{0} F_{X_0}(\bar{X}_0, \bar{X}(T)) + \left[\Gamma_{X_0}(\bar{X}_0, \bar{X}(T))\Phi\right] - \Phi(0)\right\} X_0, \quad \forall X_0 \in \mathbb{R}^d, \\
\end{cases}
\end{align*}
\]

which implies

\[
\begin{align*}
\Phi(0) &= \mathbb{E}\left\{\Phi_{0} F_{X_0}(\bar{X}_0, \bar{X}(T)) + \left[\Gamma_{X_0}(\bar{X}_0, \bar{X}(T))\Phi\right] - \Phi(0)\right\}.
\end{align*}
\]

(4.39)

\[
\begin{align*}
\mathbb{E}\left\{\Phi_{0} F_{X_0}(\bar{X}_0, \bar{X}(T)) + \left[\Gamma_{X_0}(\bar{X}_0, \bar{X}(T))\Phi\right] - \Phi(0) - \Phi(0)^\top\right\} \geq 0.
\end{align*}
\]

(4.40)
Finally, by taking $X_0 = 0$ in (4.38), using a standard argument [40], we have the following variational inequality:

$$
\langle \ddot{\Phi}(t), B(t, \bar{X}(t), v) - B(t, \bar{X}(t), \bar{v}(t)) \rangle \\
+ \langle \ddot{\Psi}(t), \Sigma(t, \bar{X}(t), v) - \Sigma(t, \bar{X}(t), \bar{v}(t)) \rangle \\
- \frac{1}{2} \left[ \Sigma(t, \bar{X}(t), v) - \Sigma(t, \bar{X}(t), \bar{v}(t)) \right]^\top P(t) \\
\cdot \left[ \Sigma(t, \bar{X}(t), v) - \Sigma(t, \bar{X}(t), \bar{v}(t)) \right] \geq 0.
$$

(4.41)

**Step 6. Finalizing the proof.** Since $\Phi_0 \neq 0$, by rescaling, we let $\Phi_0 = 1$. Then

$$
\begin{align*}
\tilde{\Phi}(t) &= - \left[ \tilde{B}_X(t)^\top \tilde{\Phi}(t) + \tilde{\Sigma}_X(t)^\top \tilde{\Psi}(t) \right] dt + \tilde{\Psi}(t) dW(t), \\
(4.42) &\quad \text{and} \\
\tilde{\Phi}(T) &= - \left\{ F_{X_T, X_0}(\bar{X}_0, \bar{X}(T)) + \left[ \tilde{\Gamma}_{X_T, X_0}(\bar{X}_0, \bar{X}(T)) \tilde{\Phi} \right] \right\},
\end{align*}
$$

and

$$
\begin{align*}
\tilde{\Phi}(t) &= - \left[ \tilde{B}_X(t)^\top \bar{\Phi}(t) + \tilde{\Sigma}_X(t)^\top \bar{\Psi}(t) \right] dt + \bar{\Psi}(t) dW(t), \\
(4.43) &\quad \text{and} \\
\tilde{\Phi}(T) &= - \left\{ F_{X_T, X_0}(\bar{X}_0, \bar{X}(T)) + \left[ \Gamma_{X_T, X_0}(\bar{X}_0, \bar{X}(T)) \bar{\Phi} \right] \right\},
\end{align*}
$$

Also,

$$
\tilde{\Phi}(0) = \mathbb{E} \left[ F_{X_0, X_0}(\bar{X}_0, \bar{X}(T))^\top + \Gamma_{X_0, X_0}(\bar{X}_0, \bar{X}(T)) \bar{\Phi} \right]
$$

and

$$
\mathbb{E} \left\{ F_{X_0, X_0}(\bar{X}_0, \bar{X}(T)) + \left[ \Gamma_{X_0, X_0}(\bar{X}_0, \bar{X}(T)) \bar{\Phi} \right] \right\}
$$

$$
- P(0) - \left[ \bar{\Phi}(0) + \bar{\Phi}(0)^\top \right] \geq 0.
$$

(4.46)

Note that

$$
F_{X_0, X_0}(\bar{X}_0, \bar{X}(T))^\top = \begin{pmatrix} 0 \\ \bar{f}_{y_0} \end{pmatrix}, \quad F_{X_T, X_0}(\bar{X}_0, \bar{X}(T))^\top = \begin{pmatrix} \bar{f}^\top_x \\ 0 \end{pmatrix}.
$$

On the other hand, since

$$
\langle \Gamma, \bar{\Phi} \rangle = \langle \bar{\Phi}_0, x_0 - \gamma(x_T, y_0) \rangle + \mathbb{E} \left( \bar{\Phi}_T, y_T - h(x_T, y_0) \right),
$$

$$
= \sum_{i=1}^n \bar{\Phi}_0^i [x_0^i - \gamma^i(x_T, y_0)] + \mathbb{E} \left( \sum_{j=1}^m \bar{\Phi}_T^j [y_T^j - h^j(x_T, y_0)] \right),
$$

(4.47)
we have
\[ \Gamma_{X_0}(X_0, X(T)) \Phi = \left( -\gamma_{y_0}^T \Phi_0 - \mathbb{E}[h_{y_0}^T \Phi_T] \right), \]
\[ \Gamma_{X_T}(X_0, X(T)) \Phi = \left( -\gamma_{x_T}^T \Phi_0 - h_{x_T}^T \Phi_T \right). \]

Let
\[ (4.47) \quad \bar{\Phi}(\cdot) = \begin{pmatrix} \eta(\cdot) \\ \xi(\cdot) \end{pmatrix}, \quad \bar{\Psi}(\cdot) = \begin{pmatrix} \zeta(\cdot) \\ \zeta(\cdot) \end{pmatrix}. \]

Then it follows from (4.42), (4.45), and (4.47) that
\[ \begin{pmatrix} \eta(0) \\ \xi(0) \end{pmatrix} = \begin{pmatrix} \Phi_0 \\ \mathbb{E}[\bar{f}_{y_0} - \gamma_{y_0}^T \Phi_0 - h_{y_0}^T \Phi_T] \end{pmatrix}, \]
\[ \begin{pmatrix} \eta(T) \\ \xi(T) \end{pmatrix} = \begin{pmatrix} -\bar{f}_{x_T} + \gamma_{x_T}^T \Phi_0 + h_{x_T}^T \Phi_T \\ -\Phi_T \end{pmatrix}. \]

Hence,
\[ \bar{\Phi} = \begin{pmatrix} \Phi_0 \\ \Phi_T \end{pmatrix} = \begin{pmatrix} \eta(0) \\ -\xi(T) \end{pmatrix}, \]
and
\[ (4.48) \quad \begin{cases} \xi(0) = \mathbb{E} \left[ h_{x_T}^T \xi(T) - \gamma_{y_0}^T \eta(0) + \bar{f}_{y_0} \right], \\
\eta(T) = -h_{x_T}^T \xi(T) + \gamma_{x_T}^T \eta(0) - \bar{f}_{x_T} \end{cases}. \]

These are mixed initial-terminal conditions in (3.1). By (4.47), we may rewrite the equations in (4.42) as
\[ d \begin{pmatrix} \eta(t) \\ \xi(t) \end{pmatrix} = - \begin{bmatrix} g_x(t)^T & \gamma_x(t)^T \\ g_y(t)^T & \gamma_y(t)^T \end{bmatrix} \begin{pmatrix} \eta(t) \\ \xi(t) \end{pmatrix} + \begin{pmatrix} \sigma_x(t)^T \\ \sigma_y(t)^T \end{pmatrix} \begin{pmatrix} \zeta(t) \\ \zeta(t) \end{pmatrix} dt + \begin{pmatrix} \zeta(t) \\ \zeta(t) \end{pmatrix} dW(t). \]

Variational inequality (4.41) takes the following form:
\[ 0 \leq \langle \eta(t), b(t, \bar{x}(t), \bar{y}(t), z, u) - b(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)) \rangle \\
+ \langle \xi(t), g(t, \bar{x}(t), \bar{y}(t), z, u) - g(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)) \rangle \\
+ \langle \zeta(t), \sigma(t, \bar{x}(t), \bar{y}(t), z, u) - \sigma(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)) \rangle + \langle \bar{\zeta}(t), z - \bar{z}(t) \rangle \\
- \frac{1}{2} \left( P(t) \begin{bmatrix} \sigma(t, \bar{x}(t), \bar{y}(t), z, u) - \sigma(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t)) \\ z - \bar{z}(t) \end{bmatrix} \right). \]

Thus, taking \( u = \bar{u}(t) \) and \( z = \bar{z}(t) + \varepsilon z_0 \), then dividing by \( \varepsilon \) and sending \( \varepsilon \to 0 \), we get
\[ 0 \leq \langle \eta(t), b_z(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t))z_0 \rangle + \langle \xi(t), g_z(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t))z_0 \rangle \\
+ \langle \zeta(t), \sigma_z(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t))z_0 \rangle + \langle \bar{\zeta}(t), z_0 \rangle \quad \forall z_0 \in \mathbb{R}^m. \]
Hence,

\[
\begin{aligned}
\tilde{\zeta}(t) = - \left[ h_z(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t))^\top \eta(t) + g_z(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t))^\top \xi(t) \\
+ \sigma_z(t, \bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{u}(t))^\top \zeta(t) \right].
\end{aligned}
\]

Then (4.50) together with (4.48) gives (3.1). Next, we have

\[
F_{X_T X_0}(\bar{X}_0, \bar{X}(T)) + \left[ \Gamma_{X_T X_0}(\bar{X}_0, \bar{X}(T)) \Phi \right]
= \begin{pmatrix}
\bar{f}_{x_T y_0} - \sum_{i=1}^n \tilde{\phi}_{0 i} \bar{\gamma}_{x_T y_0}^i - \sum_{j=1}^m \bar{\phi}_{j i} \bar{h}_{x_T y_0}^j & 0 \\
0 & 0
\end{pmatrix}
= \begin{pmatrix}
\bar{f}_{x_T y_0} - \sum_{i=1}^n \eta^i(0) \bar{\gamma}_{x_T y_0}^i + \sum_{j=1}^m \xi^j(T) \bar{h}_{x_T y_0}^j & 0 \\
0 & 0
\end{pmatrix},
\]

Thus, the terminal condition in (3.2) follows. Further,

\[
F_{X_T X_T}(\bar{X}_0, \bar{X}(T)) + \left[ \Gamma_{X_T X_T}(\bar{X}_0, \bar{X}(T)) \Phi \right]
= \begin{pmatrix}
\bar{f}_{x_T x_T} - \sum_{i=1}^n \tilde{\phi}_{0 i} \bar{\gamma}_{x_T x_T}^i - \sum_{j=1}^m \bar{\phi}_{j i} \bar{h}_{x_T x_T}^j & 0 \\
0 & 0
\end{pmatrix}
= \begin{pmatrix}
\bar{f}_{x_T x_T} - \sum_{i=1}^n \eta^i(0) \bar{\gamma}_{x_T x_T}^i + \sum_{j=1}^m \xi^j(T) \bar{h}_{x_T x_T}^j & 0 \\
0 & 0
\end{pmatrix},
\]

which leads to the terminal condition in (3.3). Similarly, we have

\[
F_{X_0 X_0}(\bar{X}_0, \bar{X}(T)) + \left[ \Gamma_{X_0 X_0}(\bar{X}_0, \bar{X}(T)) \Phi \right]
= \begin{pmatrix}
0 & 0 \\
0 & 0
\end{pmatrix}
= \begin{pmatrix}
0 & 0 \\
0 & 0
\end{pmatrix},
\]

Hence, inequality (3.5) follows from (4.46). Next, by taking \(z = \bar{z}(t)\) in the variational
and (4.51), we obtain (3.6). This completes the proof of our optimality variational principle.

In the case that
\[ b(t, x(t), y(t), z(t), u(t)) = \begin{cases} 0 & \text{if } \langle \eta(t), b(t, x(t), y(t), z(t), u(t)) \rangle \\
\frac{1}{2} \left( \begin{array}{c} \sigma(t, x(t), y(t), z(t), u(t)) - \sigma(t, x(t), y(t), z(t), \bar{u}(t)) \\
\sigma(t, x(t), y(t), z(t), u(t)) - \sigma(t, x(t), y(t), z(t), \bar{u}(t)) \end{array} \right) \end{cases} \]

which gives (3.7). Finally, in (4.50), dividing \( |z - \bar{z}(t)|^2 \) and then sending \( |z - \bar{z}(t)| \to 0 \), and making use of the second order expansion of
\[ z \to b(t, x(t), y(t), z, \bar{u}(t)), \sigma(t, x(t), y(t), z, \bar{u}(t)), g(t, x(t), y(t), z, \bar{u}(t)) \]
and (4.51), we obtain (3.6). This completes the proof of our optimality variational principle.

Note that for any \( z_0 \in \mathbb{R}^m \) with \( |z_0| = 1 \) and \( \lambda > 0 \), by taking \( z = \lambda z_0 \) in (4.49), and dividing \( \lambda^2 \), then sending \( \lambda \to \infty \), we can further obtain
\[ \lim_{\lambda \to \infty} \left( \frac{\sigma(t, x(t), y(t), z_0, \bar{u}(t))}{\lambda} \right) = 0. \]

In the case that
\[ \lim_{|z| \to 0} \frac{|\sigma(t, x, y, z, u)|}{|z|} = 0, \]
we have \( P_3(t) \leq 0, \quad t \in [0, T], \text{ a.s.} \).

5. A linear-quadratic problem. In this section, we briefly consider a linear quadratic case, which serves as an illustrative example of our main result. Consider the following controlled FBSDE:

\[
\begin{align*}
dx(t) &= \left[ b_1 x(t) + b_2 y(t) + b_3 z(t) + b_4 u(t) \right] dt \\
&\quad + \left[ \sigma_1 x(t) + \sigma_2 y(t) + \sigma_3 z(t) + \sigma_4 u(t) \right] dW(t), \\
\frac{d}{dt}y(t) &= \left[ g_1 x(t) + g_2 y(t) + g_3 z(t) + g_4 u(t) \right] dt + z(t) dW(t), \\
x(0) &= \mathbb{E}\left[ \gamma_1 x(T) \right] + \gamma_2 y(0), \quad y(T) = h_1 x(T) + h_2 y(0),
\end{align*}
\]

where \( b_i, \sigma_i, g_i, \gamma_i, h_i \) are all real constants. Consider the cost functional
\[
J(u(\cdot)) = \mathbb{E} \left[ \frac{1}{2} \left( \begin{array}{cc}
G_1 & G_2 \\
G_2 & G_3
\end{array} \right) \left( \begin{array}{c} x(T) \\
y_0
\end{array} \right) \right],
\]
with $G_1, G_2, G_3 \in \mathbb{R}$. The control domain $U \subseteq \mathbb{R}$ could be very arbitrary; in particular, $U$ does not have to be a convex set. We let

$$
\mathcal{U}[0,T] = \left\{ u(\cdot) \in L^2(\mathbb{R}; \mathbb{R}) \mid u(t) \in U, \ t \in [0,T] \right\}.
$$

Assume that

$$
\begin{pmatrix}
2g_1 & b_1 + g_2 & \sigma_1 + g_3 \\
 b_1 + g_2 & 2b_2 & \sigma_2 + b_3 \\
 \sigma_1 + g_3 & \sigma_2 + b_3 & 2\sigma_3
\end{pmatrix}
\begin{pmatrix}
g_1 \\
g_2 \\
g_3
\end{pmatrix}
\approx 0
$$

and

$$
\begin{pmatrix}
2h_1 \\
h_2 - \gamma_1 \\
-2\gamma_2
\end{pmatrix}
\gg 0,
$$

where $A \gg 0$ stands for $A$ being positive definite, and $A \ll 0$ stands for $-A \gg 0$. Then by [39], for any $u(\cdot) \in \mathcal{U}[0,T]$, there exists a unique process state $(x(\cdot), y(\cdot), z(\cdot))$. Now, suppose the corresponding optimal control problem admits an optimal 4-tuple $(\bar{x}(\cdot), \bar{y}(\cdot), \bar{z}(\cdot), \bar{u}(\cdot))$. Let us look at the optimality variational principle. The first adjoint equation reads

$$
\begin{cases}
\begin{aligned}
d\xi(t) &= -\left[ g_2 \xi(t) + b_2 \eta(t) + \sigma_2 \zeta(t) \right] dt \\
- \left[ g_3 \xi(t) + b_3 \eta(t) + \sigma_3 \zeta(t) \right] dW(t),
\end{aligned}
\end{cases}
$$

and

$$
\begin{cases}
\begin{aligned}
d\eta(t) &= -\left[ g_1 \xi(t) + b_1 \eta(t) + \sigma_1 \zeta(t) \right] dt + \zeta(t) dW(t),
\end{aligned}
\end{cases}
$$

$$
\begin{cases}
\begin{aligned}
\xi(0) &= \mathbb{E} \left[ h_2 \xi(T) \right] - \gamma_2 \eta(0) + \mathbb{E} \left[ G_2 \bar{x}(T) \right] + G_3 \bar{y}(0),
\end{aligned}
\end{cases}
$$

$$
\begin{cases}
\begin{aligned}
\eta(T) &= -h_1 \xi(T) + \gamma_1 \eta(0) - G_1 \bar{x}(T) - G_2 \bar{y}(0).
\end{aligned}
\end{cases}
$$

The BSDE for $(\Phi(\cdot), \Psi(\cdot))$ becomes

$$
\begin{cases}
\begin{aligned}
d\Phi(t) &= -\left[ B^\top \Phi(t) + \Sigma^\top \Psi(t) \right] dt + \Psi(t) dW(t),
\end{aligned}
\end{cases}
$$

\hspace{1cm} $\Phi(T) = -\begin{pmatrix} 0 \\ G_2 \\ 0 \end{pmatrix},$

and the second adjoint equation becomes

$$
\begin{cases}
\begin{aligned}
dP(t) &= -\left[ B^\top P(t) + P(t) B + \Sigma^\top P(t) \Sigma + \Sigma^\top Q(t) + Q(t) \Sigma \right] dt \\
+ Q(t) dW(t),
\end{aligned}
\end{cases}
$$

\hspace{1cm} $P(T) = -\begin{pmatrix} G_1 \\ 0 \\ 0 \end{pmatrix},$

where

$$
B = \begin{pmatrix} b_1 & b_2 \\ g_1 & g_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1 & \sigma_2 \\ 0 & 0 \end{pmatrix}.
$$

Note that since all the coefficients in (5.6)–(5.7) are constants, and the terminal conditions are deterministic, we must have

$$
\Psi(\cdot) = 0, \quad Q(\cdot) = 0,
$$
and $\Phi(\cdot)$ and $P(\cdot)$ are deterministic. Consequently,

\begin{equation}
\dot{\Phi}(t) = -B^T \Phi(t), \quad \Phi(T) = -\begin{pmatrix} 0 & 0 \\ G_2 & 0 \end{pmatrix}
\end{equation}

and

\begin{equation}
\dot{P}(t) = -\left[ B^T P(t) + P(t)B + \Sigma^T P(t) \Sigma \right], \quad P(T) = -\begin{pmatrix} G_1 & 0 \\ 0 & 0 \end{pmatrix}.
\end{equation}

Clearly, the above two are terminal value problems for linear ODEs which only depend on the coefficients of the state equation and the cost functional, and they are independent of the optimal 4-tuple $(\bar{x}(\cdot), \bar{y}(\cdot), \bar{z}(\cdot), \bar{u}(\cdot))$. Hence, conditions (3.5)–(3.6), which take the form

\begin{equation}
\begin{cases}
(0 \\ 0 \\ G_3 \end{pmatrix} - P(0) - \left[ \Phi(0) + \Phi(0)^T \right] \geq 0, \\
(\sigma_3 I \begin{pmatrix} \sigma_3 \\ I \end{pmatrix} \right) P(t) \begin{pmatrix} \sigma_3 \\ I \end{pmatrix} \leq 0 \quad \text{a.e. } t \in [0, T], \text{ a.s.}
\end{cases}
\end{equation}

actually present necessary conditions for the coefficients of the state equation and the cost functional in order that the corresponding Problem (C) admits an optimal 4-tuple. The optimality variational inequality (3.7) now takes the form

\begin{equation}
\left[ g_4 \xi(t) + b_4 \eta(t) + \sigma_4 \zeta(t) \right] \left[ u - \bar{u}(t) \right] - \frac{1}{2} P_1(t) \sigma_4^2 \left[ u - \bar{u}(t) \right]^2 \geq 0 \quad \forall u \in U.
\end{equation}

This implies that

\begin{equation}
\begin{cases}
g_4 \xi(t) + b_4 \eta(t) + \sigma_4 \zeta(t) \geq \frac{1}{2} P_1(t) \sigma_4^2 \left[ u - \bar{u}(t) \right] \quad \forall u \in U, \ u > \bar{u}(t), \\
g_4 \xi(t) + b_4 \eta(t) + \sigma_4 \zeta(t) \leq \frac{1}{2} P_1(t) \sigma_4^2 \left[ u - \bar{u}(t) \right] \quad \forall u \in U, \ u < \bar{u}(t).
\end{cases}
\end{equation}

Now, let us look at an interesting special case. Suppose

\begin{equation}
b_2 = \sigma_2 = 0.
\end{equation}

Note that the above conditions do not mean that the state equation (the FBSDE) is decoupled. Although under the above conditions, condition (5.3) fails, it is still possible that the so-called partial-monotone condition is satisfied, and together with other suitable conditions, a Lyapunov operator will exist so that (H0) holds [39]. Now, under (5.13),

\begin{equation}
e^{B_t} = \begin{pmatrix} e^{b_1 t} \\ \beta(t) \end{pmatrix}, \quad \beta(t) = \begin{cases}
g_1 \left( \frac{e^{b_1 t}}{b_1} - 1 \right), & b_1 \neq 0, \\
g_1 t, & b_1 = 0.
\end{cases}
\end{equation}

Hence,

\begin{equation}
\Phi(t) = -\begin{pmatrix} 0 & 0 \\ G_2 & 0 \end{pmatrix}, \quad t \geq 0.
\end{equation}
Also, under (5.13), (5.9) becomes

\[
\begin{pmatrix}
\dot{P}_1(t) & \dot{P}_2(t) \\
\dot{P}_2(t) & \dot{P}_3(t)
\end{pmatrix}
= -\begin{bmatrix}
b_1 & g_1 \\
0 & 2g_2
\end{bmatrix}
\begin{pmatrix}
P_1(t) & P_2(t) \\
P_2(t) & P_3(t)
\end{pmatrix}
+ \begin{bmatrix}
\sigma_1 & 0 \\
0 & 0
\end{bmatrix}
\begin{pmatrix}
P_1(t) & P_2(t) \\
P_2(t) & P_3(t)
\end{pmatrix}
\begin{bmatrix}
b_1 & 0 \\
g_1 & g_2
\end{bmatrix}

= -\begin{bmatrix}
(2b_1 + \sigma_1^2)P_1(t) + 2g_1P_2(t) \\
(b_1 + g_2)P_2(t) + g_1P_3(t)
\end{bmatrix}
\begin{pmatrix}
b_1 & 0 \\
g_1 & g_2
\end{bmatrix}.
\]

Thus, together with the terminal condition, we see that

\[
P_2(\cdot) = P_3(\cdot) = 0.
\]

Further,

\[
\dot{P}_1(t) = -(2b_1 + \sigma_1^2)P_1(t), \quad P_1(T) = -G_1,
\]

which leads to

\[
P_1(t) = -e^{(2b_1 + \sigma_1^2)(T-t)}G_1.
\]

Consequently, (5.10) becomes

\[
\begin{cases}
0 \leq \begin{pmatrix} 0 & 0 \\ 0 & G_3 \end{pmatrix} - P(0) - \begin{bmatrix} \Phi(0) + \Phi(0)^T \end{bmatrix} \\
e^{(2b_1 + \sigma_1^2)T}G_1 G_2, \\
0 \geq \begin{pmatrix} \sigma_3 \\ I \end{pmatrix} P(t) \begin{pmatrix} \sigma_3 \\ I \end{pmatrix} = -\sigma_3^2 e^{(2b_1 + \sigma_1^2)(T-t)}G_1 \\
\text{a.e. } t \in [0, T], \text{ a.s.}
\end{cases}
\]

Thus, it is necessary that

\[
G_1, G_3 \geq 0, \quad e^{(2b_1 + \sigma_1^2)T}G_1 G_3 - G_2^2 \geq 0.
\]

Note that for a linear-quadratic optimal control problem of (forward) SDEs, with all the weight matrices in the integrand of the cost functional being zero, according to [16], as long as the weight matrix in the terminal state penalty term is positive definite, the corresponding linear-quadratic problem is well-posed (meaning that optimal control exists uniquely). From this, it seems to be natural that our current optimal control problem should admit an optimal 4-tuple if we assume condition

\[
\begin{pmatrix} G_1 & G_2 \\ G_2 & G_3 \end{pmatrix} \succ 0,
\]

which is equivalent to

\[
G_1, G_3 > 0, \quad G_1 G_3 - G_2^2 > 0.
\]
But, (5.20) does not necessarily imply the second relation in (5.18) when $2b_1 + \sigma_1^2 < 0$. This shows that even for a linear-quadratic optimal control problem of FBSDEs (with mixed initial-terminal conditions), there are some interesting things to be explored. We will carefully discuss some relevant issues in a forthcoming paper.

On the other hand, for the current case, condition (5.12) becomes

$$
\begin{aligned}
&\begin{cases}
  g_4 \zeta(t) + b_4 \eta(t) + \sigma_4 \zeta(t) \geq -\frac{1}{2} e^{(2b_1 + \sigma_1^2)(T-t)} G_1 \sigma_4^2 [u - \bar{u}(t)] \\
  &\quad \forall u \in U, u > \bar{u}(t), \\
  g_4 \zeta(t) + b_4 \eta(t) + \sigma_4 \zeta(t) \leq -\frac{1}{2} e^{(2b_1 + \sigma_1^2)(T-t)} G_1 \sigma_4^2 [u - \bar{u}(t)] \\
  &\quad \forall u \in U, u < \bar{u}(t).
\end{cases}
\end{aligned}
$$

(5.21)

Hence, in the case that $\sigma_4 = 0$, the above leads to

$$
(5.22) \quad g_4 \zeta(t) + b_4 \eta(t) + \sigma_4 \zeta(t) = 0, \quad t \in [0, T], \text{ a.s.}
$$

and in the case that $\sigma_4 \neq 0$, relations (5.21) imply that

$$
(5.23) \quad \bar{u}(t) - \inf_{u \in U, u > \bar{u}(t)} u \leq \frac{2 [g_4 \zeta(t) + b_4 \eta(t) + \sigma_4 \zeta(t)]}{e^{(2b_1 + \sigma_1^2)(T-t)} G_1 \sigma_4^2} \leq \bar{u}(t) - \sup_{u \in U, u < \bar{u}(t)} u.
$$

Note that when $U$ is discrete and not a singleton, the leftmost and the rightmost sides of the above are different. Finally, when $U = \mathbb{R}$, we have (5.22), regardless of either $\sigma_4 = 0$ or $\sigma_4 \neq 0$. This gives a necessary condition for the optimal 4-tuple $(\bar{x}(\cdot), \bar{y}(\cdot), \bar{z}(\cdot), \bar{u}(\cdot))$.

Acknowledgment. The author thanks the anonymous referee for his/her careful reading of the previous version of the paper.

REFERENCES


