SCIENCE CHINA
Mathematics

• ARTICLES •

June 2015 Vol. 58 No. 6: 1265–1284
doi: 10.1007/s11425-015-4972-x

Maximum principle for optimal control of neutral stochastic functional differential systems

WEI WenNing

School of Mathematical Sciences, Fudan University, Shanghai 200433, China
Email: wnwei@fudan.edu.cn

Received October 19, 2012; accepted October 3, 2014; published online January 23, 2015

Abstract This paper is concerned with optimal control of neutral stochastic functional differential equations (NSFDEs). The Pontryagin maximum principle is proved for optimal control, where the adjoint equation is a linear neutral backward stochastic functional equation of Volterra type (VNBSFE). The existence and uniqueness of the solution are proved for the general nonlinear VNBSFEs. Under the convexity assumption of the Hamiltonian function, a sufficient condition for the optimality is addressed as well.

Keywords neutral stochastic functional differential equation, neutral backward stochastic functional equation of Volterra type, stochastic optimal control, Pontryagin maximum principle

MSC(2010) 93E20, 60H20

Citation: Wei W N. Maximum principle for optimal control of neutral stochastic functional differential systems. Sci China Math, 2015, 58: 1265–1284, doi: 10.1007/s11425-015-4972-x

1 Introduction

In this paper, we consider the following stochastic optimal control problem:

$$\min_{u\in\mathcal{U}_{ad}} E\bigg[\int_0^T l(t,X^t,u(t))dt\bigg], \tag{1.1}$$

subject to the neutral stochastic functional differential equation (NSFDE),

$$\begin{cases} d[X(t)-g(t,X^t,u(t))]=b(t,X^t,u(t))dt+\sigma(t,X^t,u(t))dW(t), \quad t\in[0,T], \\ X(t)=\phi(t), \quad t\in[-\delta,0], \end{cases} \tag{1.2}$$

where $\delta \geqslant 0$ is a constant, X^t denotes the restricted path of X on $[t-\delta,t]$, $u(\cdot)$ is the control process, W is a d-dimensional Brownian motion, g,b and σ are suitable functionals on $\Omega\times[0,T]\times C([-\delta,0],\mathbb{R}^n)\times\mathbb{R}^m$, and ϕ is a continuous function on $[-\delta,0]$. We establish the Pontryagin maximum principle by introducing the neutral backward stochastic functional equations of Volterra type (VNBSFEs) as the adjoint equations.

The difficulty in establishing the maximum principle mainly relies on the construction and resolution to the adjoint equation. The solution to NSFDE might not be a semi-martingale due to the part $g(\cdot,\cdot)$ in the left-hand side. Therefore, the traditional method (see, e.g., [4–6]) for the optimal control problem of stochastic differential equations (SDEs) would not work directly. To interpret the adjoint equations, we derive a class of neutral backward stochastic functional equations of Volterra type (VNBSFEs) which, generally, is of the form

$$\begin{cases} Y(t) - G(t, Y_t) = \Psi(t) + \int_t^T f(t, s, Y_s, Z(t, s), Z(s, t; \delta))ds - \int_t^T Z(t, s)dW(s), \quad t \in [0, T], \\ Y(t) = \xi(t), \quad t \in (T, T + \delta], \end{cases} \tag{1.3}$$

where Y_t represents the restricted path of Y on $[t, t + \delta]$, and $Z(s, t; \delta)$ denotes the restriction of Z on $[s, s + \delta] \times [t, t + \delta]$. To the best of our knowledge, such a VNBSFE (1.3) is new and we prove the uniqueness and existence of the solution in Section 3. It is worth noting that Wei [34] considered a special case when $\Psi(t) \equiv \Psi$ and $f(t, s, \cdot, \cdot, \cdot) = f(s, \cdot, \cdot, 0)$, proved the existence and uniqueness of the solution and established a maximum principle for optimal control with state processes being driven by a special class of NSFDEs. In addition, when $G \equiv 0$ and $\delta = 0$, VNBSFE (1.3) reads

$$Y(t) = \Psi(t) + \int_t^T f(t, s, Y(s), Z(t, s), Z(s, t))ds - \int_t^T Z(t, s)dW(s),$$

which becomes the so-called backward stochastic Volterra integral equation (BSVIE, see Yong [35]).

With the fixed point method, we first prove the existence and uniqueness of the M-solution (see Definition 3.1 below) to VNBSFE (1.3). On basis of the well-posedness of the linear VNBSFEs and the dual analysis between linear NSFDEs and VNBSFEs, the maximum principle is then established for the optimal control problem (1.1). A sufficient condition for optimal control is also derived under the convexity assumption of Hamiltonian function. When the state equation is reduced to a stochastic differential equation (SDE), the maximum principle herein is bridged to the traditional one by Bismut [6]. We compare these two maximum principles and show the explicit relations.

In NSFDE (1.2), the evolution rate of state depends not only on the present state, but also on the past state and the past evolution rate of the state. This kind of equation models a large class of systems with after-effect, which is widely used in biology, mechanics, physics, medicine and economics, such as population sizes, commodity supply fluctuations and so on. See [9, 13, 14, 20–22] and reference therein. Many research works on NSFDEs focus on the well-posedness and stability of the solutions, see [17, 23, 26, 30, 32, 33] and reference therein. While the optimal control problem of deterministic neutral functional differential equation has been extensively discussed (see [3, 15, 18, 19, 31]), the stochastic case has just caused attentions recently (see [2, 25, 27]). A particular stochastic optimal control problem with time delay was studied by Øksendal et al. [11, 27] where $g(\cdot, X^{\cdot}, u(\cdot)) \equiv \gamma(\cdot)$ and as a nondecreasing process, $\gamma(\cdot)$ is modeled as operations to change the fish population in agriculture or the value of an investment in finance. As an infinite dimensional counterpart of (1.1), the linear quadratic stochastic optimal control problem of neutral type was considered by Liu [25], where the diffusion coefficient $\sigma(t, X^t, u(t))$ is a constant and is not controlled. Ahmed [2] considered the optimal control problems driven by a class of second order stochastic neutral differential equations on Banach spaces with the drifts depending linearly on the control variable and the diffusion term uncontrolled $\sigma(t, X^t, u(t)) \equiv \mathcal{D}(t)$ (see [2, (14) and (24)]). It is worth noting that the (controlled) neutral stochastic differential equations of Ahmed [2] can be seen as Hilbert space-valued SDEs, which are essentially different from NSFDE (1.2) and systems by Liu [25]. We also note that in [7, 8, 16, 36], the maximum principle was established for controlled systems driven by stochastic functional differential equations which are not of neutral type. In this work, we shall extend the previous theory to study the stochastic optimal control problem for the general neutral stochastic functional differential equations (1.2), which seems to be new and is quite important from both theoretic and practical viewpoints.

The rest of this paper is organized as follows. In Section 2, we give some notations and introduce the optimal control problem. In Section 3, we are concerned with the well-posedness of VNBSFE (1.3) and prove the existence and uniqueness of the M-solution. Section 4 is devoted to the duality of linear NSFDEs and VNBSFEs. In Section 5, we establish the maximum principle for controlled NSFDE (1.2) with Lagrange type cost functional. Finally, in Section 6, two related topics are discussed: We first consider the particular case where the state equation is reduced to an SDE associated with $g \equiv 0$ and $\delta = 0$ in (1.2), and compare our maximum principle herein with the traditional one of [6], and address the explicit relations between them; in the second subsection, we give a sufficient condition for optimal control when Hamiltonian function is convex, and a simple linear quadratic (LQ) problem is studied as an application.

2 Preliminaries

In this section, we shall give some notation and introduce the stochastic optimal control problem with the state process driven by NSFDEs.

Throughout this work, $\delta \geqslant 0$ and $T > 0$ are two universal constants. Let $(\Omega, \mathscr{F}, \mathbb{F}, P)$ be a complete filtered probability space on which a d-dimensional Brownian motion $W = \{W(t) : t \in [0, T+\delta]\}$ is defined. $\{\mathscr{F}_t : t \in [0, T+\delta]\}$ is the natural filtration of W augmented by all P-null sets in $\mathscr{F}$. Define $\mathscr{F}_t := \mathscr{F}_0$ for any $t \in [-\delta, 0]$. Then $\mathbb{F} := \{\mathscr{F}_t : t \in [-\delta, T+\delta]\}$ is a filtration satisfying the usual conditions.

2.1 Notation

For each vector or matrix A, denote by A' the transpose of A. Denote by H some Euclidean space, such as $\mathbb{R}^n, \mathbb{R}^{n\times d}$, etc., and by $|\cdot|$ and $\langle \cdot, \cdot \rangle$ the norm and the inner product in H, respectively. In particular, for any $A, B \in \mathbb{R}^{n\times d}$, $\langle A, B\rangle = \mathrm{tr}(A'B)$. Define

$$\mathbb{L}^2(\Omega; H) := \{\eta : \Omega \to H,\ \mathscr{F}_T\text{-measurable} \mid E[|\eta|^2] < +\infty\}.$$

For $r, s, \tau, \nu \in [-\delta, T+\delta]$, $r \leqslant s$ and $\tau \leqslant \nu$, define

$$\begin{aligned}
\mathscr{L}^2_{\mathbb{F}}(r, s; H) &:= \left\{\theta : [r, s] \times \Omega \to H,\ \mathbb{F}\text{-adapted} \,\middle|\, E\int_r^s |\theta(u)|^2 du < +\infty\right\},\\
\mathscr{S}^2_{\mathbb{F}}([r, s]; H) &:= \left\{\theta : [r, s] \times \Omega \to H,\ \mathbb{F}\text{-adapted, path-continuous} \,\middle|\, E\sup_{r\leqslant u\leqslant s} |\theta(u)|^2 < +\infty\right\},\\
\mathbb{L}^2(r, s; \mathbb{L}^2(\Omega; H)) &:= \left\{\psi : [r, s] \times \Omega \to H,\ \mathscr{B}([r, s]) \times \mathscr{F}_T\text{-measurable} \,\middle|\, E\int_r^s |\psi(u)|^2 du < +\infty\right\},\\
\mathbb{L}^2(r, s; \mathbb{L}^2(\tau, \nu; H)) &:= \left\{v : [r, s] \times [\tau, \nu] \to H,\ \text{jointly-measurable} \,\middle|\, \int_r^s\int_\tau^\nu |v(t, s)|^2 ds dt < +\infty\right\},\\
\mathbb{L}^2(r, s; \mathscr{L}^2_{\mathbb{F}}(\tau, \nu; H)) &:= \Big\{\vartheta : [r, s] \times [\tau, \nu] \times \Omega \to H \text{ is } \mathscr{B}([r, s] \times [\tau, \nu]) \times \mathscr{F}_T\text{-measurable},\\
&\qquad \vartheta(t, \cdot) \text{ is } \mathbb{F}\text{-adapted for all } t \in [r, s] \,\Big|\, E\int_r^s\int_\tau^\nu |\vartheta(t, s)|^2 ds dt < +\infty\Big\}.
\end{aligned}$$

For simplicity, denote

$$\mathscr{H}^2(r, s) := \mathscr{L}^2_{\mathbb{F}}(r, s; \mathbb{R}^n) \times \mathbb{L}^2(r, s; \mathscr{L}^2_{\mathbb{F}}(r, s; \mathbb{R}^{n\times d})),$$

equipped with norm

$$\|(\theta, \vartheta)\|^2_{\mathscr{H}^2(r,s)} = E\left[\int_r^s |\theta(u)|^2 du + \int_r^s\int_r^s |\vartheta(\nu, u)|^2 du d\nu\right].$$

Finally, for simplicity, we set

$$E_s[\cdot] = E[\cdot \mid \mathscr{F}_s], \quad s \in [0, T].$$

2.2 The optimal control problem

Consider the following controlled NSFDE:

$$\begin{cases} d[X(t)-g(t,X^t)] = b(t,X^t,u(t))dt + \sigma(t,X^t,u(t))dW(t), & t\in[0,T], \\ X(t)=\phi(t), \quad t\in[-\delta,0], \end{cases} \tag{2.1}$$

where X^t denotes the restricted path of X on $[t-\delta,t]$, $\phi\in\mathscr{S}^2_{\mathbb{F}}([-\delta,0];\mathbb{R}^n)$,

$$\begin{aligned} &g:[0,T]\times\Omega\times C([-\delta,0];\mathbb{R}^n)\to\mathbb{R}^n,\\ &b:[0,T]\times\Omega\times C([-\delta,0];\mathbb{R}^n)\times\mathbb{R}^m\to\mathbb{R}^n,\\ &\sigma:[0,T]\times\Omega\times C([-\delta,0];\mathbb{R}^n)\times\mathbb{R}^m\to\mathbb{R}^{n\times d}, \end{aligned}$$

are jointly measurable, and $g(\cdot,\psi)$, $b(\cdot,\psi,u)$ and $\sigma(\cdot,\psi,u)$ are $\mathbb{F}$-progressively measurable for any

$$(\psi,u)\in C([-\delta,0];\mathbb{R}^n)\times\mathbb{R}^m.$$

For simplicity, we only discuss the case where g does not depend on the control u. For the case where g depends on u, see Remark 5.1 below.

Let $U\subseteq\mathbb{R}^m$ be a nonempty convex set. Denote the admissible control set by

$$\mathcal{U}_{ad}:=\left\{u(\cdot):[0,T]\times\Omega\to U,\ \mathbb{F}\text{-progressively measurable}\ \middle|\ E\int_0^T|u(t)|^2dt<+\infty\right\}.$$

For any $u(\cdot)\in\mathcal{U}_{ad}$, we consider the following cost functional:

$$J(u(\cdot))=E\left[\int_0^T l(t,X^t,u(t))dt\right],$$

where

$$l:[0,T]\times\Omega\times C([-\delta,0];\mathbb{R}^n)\times U\to\mathbb{R}$$

is jointly measurable, and $l(\cdot,\psi,u)$ is $\mathbb{F}$-progressively measurable for any $(\psi,u)\in C([-\delta,0];\mathbb{R}^n)\times U$. $\forall\,u(\cdot)\in\mathcal{U}_{ad}$, denote by $X(\cdot)$ the solution to NSFDE (2.1) associated with $u(\cdot)$, then $(X(\cdot),u(\cdot))$ is called the admissible pair.

Our optimal control problem is to find a control $\bar{u}(\cdot)\in\mathcal{U}_{ad}$, such that

$$J(\bar{u}(\cdot))=\inf_{u(\cdot)\in\mathcal{U}_{ad}}J(u(\cdot)).$$

Denote by $\bar{X}(\cdot)$ the solution to NSFDE (2.1) associated with control $\bar{u}(\cdot)$. Then $(\bar{X}(\cdot),\bar{u}(\cdot))$ is called the optimal pair.

We introduce the following standing assumptions on the random coefficients.

(A1) For every $(\omega,t)\in\Omega\times[0,T]$, g,b,σ,l are continuously Fréchet differentiable with respect to x, and b,σ,l are continuously differentiable with respect to u, and further, the derivatives $b_x,b_u,\sigma_x,\sigma_u$ and g_x are all bounded. The derivatives l_x and l_u satisfy the linear growth, i.e., there exists some positive constant L such that for every $(\omega,t)\in\Omega\times[0,T]$,

$$|l_x(t,\psi,u)|+|l_u(t,\psi,u)|\leqslant L\Big(1+\sup_{s\in[-\delta,0]}|\psi(s)|+|u|\Big),\quad\forall\,(\psi,u)\in([-\delta,0];\mathbb{R}^n)\times U.$$

(A2) $b(\cdot,0,0),\sigma(\cdot,0,0),l(\cdot,0,0)\in\mathscr{L}^2_{\mathbb{F}}(0,T;H)$, $H=\mathbb{R}^n,\mathbb{R}^{n\times d},\mathbb{R}$, respectively. For every $\omega\in\Omega$, both g and g_x are continuous in t, and there is a constant $0<\kappa<1$, such that $\|g_x\|\leqslant\kappa$.

Remark 2.1. Here, $\kappa<1$ is assumed to avoid the degenerate case like $g(t,X^t)\equiv X(t)$ and to ensure the existence and uniqueness of the solution to NSFDE (2.1) (see, e.g., [26]).

By Assumption (A1), for each $(\omega,t,u)\in\Omega\times[0,T]\times U$, $(g,b,\sigma,l)(\omega,t,\cdot,u)$ can be seen as some finite dimensional space (denoted by H) valued functional of the form

$$\mathcal{M}:\, C([-\delta,0];\mathbb{R}^n)\to H,$$

and the continuous Fréchet derivative $\mathcal{M}_x$ exists and is continuous. In fact, the Fréchet derivative $\mathcal{M}_x(\cdot)$ is a linear operator from $C([-\delta,0];\mathbb{R}^n)$ to H. By the Riesz representation theorem, we have the following lemma.

Lemma 2.1. *For each $(\omega,t,X,u)\in\Omega\times[0,T]\times C([-\delta,0];\mathbb{R}^n)\times U$, there exist $\tilde{G}(t,X,\cdot)$, $\tilde{B}(t,X,u,\cdot)$, $\tilde{\varsigma}_i(t,X,u,\cdot)\in V_0([-\delta,0];\mathbb{R}^{n\times n})$ $(i=1,\cdot,d)$ and $\tilde{L}(t,X,u,\cdot)\in V_0([-\delta,0];\mathbb{R}^{1\times n})$, such that for all $\psi\in C([-\delta,0];\mathbb{R}^n)$,*

$$g_x(t,X)\psi=\int_{-\delta}^{0}\tilde{G}(t,X,dr)\psi(r),\quad b_x(t,X,u)\psi=\int_{-\delta}^{0}\tilde{B}(t,X,u,dr)\psi(r),$$
$$\sigma_x^i(t,X,u)\psi=\int_{-\delta}^{0}\tilde{\varsigma}_i(t,X,u,dr)\psi(r),\quad l_x(t,X,u)\psi=\int_{-\delta}^{0}\tilde{L}(t,X,u,dr)\psi(r),$$

where ω is omitted and

$$V_0([-\delta,0];H):=\{f:[-\delta,0]\to H \text{ is bounded variational and left continuous on } [-\delta,0)\}.$$

Remark 2.2. Through standard arguments (for example, see [21, 26]), we conclude that under Assumptions (A1) and (A2), for any $\phi(\cdot)\in\mathscr{S}^2_{\mathbb{F}}([-\delta,0];\mathbb{R}^n)$ and $u(\cdot)\in\mathcal{U}_{ad}$, NSFDE (2.1) admits a unique solution $X(\cdot)\in\mathscr{S}^2_{\mathbb{F}}([-\delta,T];\mathbb{R}^n)$. Thus the cost function $J(u(\cdot))$ is well-defined. We note that each element $h\in\mathscr{S}^2_{\mathbb{F}}([-\delta,T];\mathbb{R}^n)$ can be seen as a $C([-\delta,0];\mathbb{R}^n)$-valued process $\{h(t+\cdot);t\in[0,T]\}$. On the other hand, if $(\phi,b,\sigma)\equiv 0$ and $g(t,\cdot)\equiv|M_t|^\alpha$ with $\alpha\in(0,1/2)$, $M_0=0$ and M_t being a continuous martingale, then (2.1) admits a unique solution $X_t=|M_t|^\alpha$ which is not a semimartingale unless $M\equiv 0$ (see [12, p. 7] and [29, Theorem 52]). Therefore, our controlled state processes can be beyond the scope of semimartingales and is essentially different from that of [2].

On the basis of Assumptions (A1) and (A2), we assume further

(A3) There exist probability measures λ_i $(i=0,1,2,3)$ on $[-\delta,0]$, and continuous functions $\bar{G}(t,X,r)$, $\bar{B}(t,X,u,r)$, $\bar{\varsigma}(t,X,u,r)$, $\bar{L}(t,X,u,r)$, such that, for all $\psi\in C([-\delta,0];\mathbb{R}^n)$ and every $(\omega,t,X,u)\in\Omega\times[0,T]\times C([-\delta,0];\mathbb{R}^n)\times U$,

$$g_x(t,X)\psi=\int_{-\delta}^{0}\bar{G}(t,X,r)\psi(r)\lambda_0(dr),\quad b_x(t,X,u)\psi=\int_{-\delta}^{0}\bar{B}(t,X,u,r)\psi(r)\lambda_1(dr),$$
$$\sigma_x(t,X,u)\psi=\int_{-\delta}^{0}\bar{\varsigma}(t,X,u,r)\psi(r)\lambda_2(dr),\quad l_x(t,X,u)\psi=\int_{-\delta}^{0}\bar{L}(t,X,u,r)\psi(r)\lambda_3(dr).$$

Remark 2.3. In Assumption (A3), we ensure the existence of both the processes $(\bar{G},\bar{B},\bar{\varsigma},\bar{L})$ and the probability measures on $([-\delta,0],\mathcal{B}([-\delta,0]))$, such that the vector valued measures admit Radon-Nikodym derivatives with respect to certain probability measures. This assumption will help us to give an explicit dual analysis between NSFDEs and VNBSFEs in Sections 4 and 5 below. In this work, (A3) is necessary to derive directly the adjoint system (4.3). One particular case for Assumption (A3) is of the linear form, i.e., in NSFDE (2.1),

$$g(t,X^t)=\hat{g}\bigg(t,\int_{-\delta}^{0}\alpha(t,r)X(t-r)\lambda_0(dr)\bigg),$$

and b,σ,l possess similar forms.

Remark 2.4. Let M be a continuous martingale with $0<|M_t|<1/2$. For $\delta>0$ and $\alpha\in(0,1/2)$, put

$$g(t,X)=|M_t|^\alpha\int_{-\delta}^{0}X(r)\,\delta^{-1}dr+|M_t|^\alpha,\quad b(t,X,u)=|M_t|^\alpha\int_{-\delta}^{0}X(r)\,\delta^{-1}dr+c_1u(t),$$

$$\sigma(t,X,u)=|M_t|^{\alpha}\int_{-\delta}^{0}X(r)\,\delta^{-1}dr+c_2u(t),\quad l(t,X,u)=|M_t|^{\alpha}\int_{-\delta}^{0}X(r)\,\delta^{-1}dr+c_3u(t)$$

with c_1, c_2 and c_3 being positive constants. This is a simple but nontrivial example for the optimal control problem above under Assumptions (A1)–(A3).

3 Well-posedness of VNBSFEs

This section is concerned with the well-posedness of neutral backward stochastic functional equations of Volterra type (VNBSFEs), which, as the adjoint equations, arise naturally from the Pontryagin maximum principle for the optimal control problem of NSFDEs (see Proposition 4.1 and Theorem 5.2 below). Inspired by Yong [35] on backward stochastic Volterra integral equations (BSVIEs), we shall give the definition of M-solution to VNBSFEs and prove the existence and uniqueness of the M-solutions to NSFDEs. In a similar way to Remark 2.2, it follows that the solutions to VNBSFEs can be beyond the scope of semimartingales. Moreover, VNBSFE is new and thus, our study herein is of independent interests.

Consider the following VNBSFE:

$$\begin{cases}Y(t)-G(t,Y_t)=\Psi(t)+\displaystyle\int_t^T f(t,s,Y_s,Z(t,s),Z(s,t;\delta))ds+\int_t^T Z(t,s)dW(s),\quad t\in[0,T],\\ Y(t)=\xi(t),\quad t\in(T,T+\delta],\end{cases}\tag{3.1}$$

where Y_t denotes the restriction of Y on $[t,t+\delta]$, $Z(\cdot,\cdot)$ is an unknown function defined on $[0,T+\delta]\times[0,T+\delta]$, $Z(t,s)$ denotes the value of Z at (t,s), and $Z(s,t;\delta)$ denotes the restriction of $Z(\cdot,\cdot)$ on $[s,s+\delta]\times[t,t+\delta]$.

For any $0\leqslant R\leqslant S\leqslant T+\delta$, define

$$\Delta[R,S]:=\{(t,s)\in[R,S]\times[R,S]\mid R\leqslant t\leqslant s\leqslant S\},$$
$$\Delta^c[R,S]:=[R,S]\times[R,S]\setminus\Delta[R,S].$$

For simplicity, denote $\Delta:=\Delta[0,T]$, $\Delta^c:=\Delta^c[0,T]$, $\Delta_\delta:=\Delta[0,T+\delta]$ and $\Delta_\delta^c:=\Delta^c[0,T+\delta]$.

(G,f) in (3.1) is called the generator of VNBSFEs. For (G,f), there exist two functionals J and F,

• $J:[0,T]\times\Omega\times\mathbb{L}^2(0,\delta;\mathbb{R}^n)\to\mathbb{R}^n$ is jointly measurable, and $J(\cdot,\phi)$ is $\mathbb{F}$-progressively measurable for any $\phi\in\mathbb{L}^2(0,\delta;\mathbb{R}^n)$;

• $F:\Delta\times\Omega\times\mathbb{L}^2(0,\delta;\mathbb{R}^n)\times\mathbb{R}^{n\times d}\times\mathbb{L}^2(0,\delta;\mathbb{L}^2(0,\delta;\mathbb{R}^{n\times d}))\to\mathbb{R}^n$ is jointly measurable, and $F(t,\cdot,\phi,z,\varphi)$ is $\mathbb{F}$-progressively measurable for all (t,ϕ,z,φ) fixed in corresponding space, and (J,F) satisfies:

(H1) There are $\kappa\in[0,1)$ and ϱ_0 being a probability measure on $[0,\delta]$, such that for any $\phi,\bar\phi\in\mathbb{L}^2(0,\delta;\mathbb{R}^n)$,

$$|J(t,\phi)-J(t,\bar\phi)|^2\leqslant\kappa\int_0^\delta|\phi(u)-\bar\phi(u)|^2\varrho_0(du).\tag{3.2}$$

(H2) There are $L>0$, ϱ_1 and ϱ_2 being probability measures on $[0,\delta]$, such that for all $(\phi,z,\varphi),(\bar\phi,\bar z,\bar\varphi)$ in corresponding space and $(t,s)\in\Delta$,

$$\begin{aligned}&|F(t,s,\phi,z,\varphi)-F(t,s,\bar\phi,\bar z,\bar\varphi)|\\&\quad\leqslant L\bigg[\int_0^\delta|\phi(u)-\bar\phi(u)|\varrho_1(du)+|w-\bar w|+\int_0^\delta|\varphi(u,u)-\bar\varphi(u,u)|\varrho_2(du)\bigg].\end{aligned}\tag{3.3}$$

(H3) (G,f) are the functionals defined

$$G(t,y_t)=E_t[J(t,y_t)],\quad f(t,s,y_s,z(t,s),z(s,t;\delta))=E_s[F(t,s,y_s,z(t,s),z(s,t;\delta))],$$

for all $(y(\cdot),z(\cdot))\in\mathscr{H}^2(0,T+\delta)$, $(t,s)\in\Delta$, $|G(\cdot,0)|\in\mathscr{L}^2_{\mathbb{F}}(0,T;\mathbb{R}^n)$, $f_0(\cdot,\cdot):=f(\cdot,\cdot,0,0,0,0)\in\mathbb{L}^2(0,T;\mathscr{L}^2_{\mathbb{F}}(0,T;\mathbb{R}^n))$.

Definition 3.1. A pair of process $(Y,Z)\in\mathscr{H}^2(0,T+\delta)$ is called an M-solution to VNBSFE (3.1), if (3.1) holds in Itô's sense for almost all $t\in[0,T+\delta]$,

$$Y(t)=E[Y(t)]+\int_0^t Z(t,s)dW(s),\quad \text{a.e. } t\in[0,T+\delta], \tag{3.4}$$

and $Z(t,s)=0$ on $(t,s)\in\Delta_\delta\setminus\Delta$.

Remark 3.1. In VNBSFE (3.1), we only set the terminal condition $Y(t)=\xi(t)$ on $(T,T+\delta]$. The value of Y at T is determined via

$$Y(T)+G(T,Y_T)=\Psi(T).$$

Compared with the classical BSDE theory (see, e.g., [28]), $Z(t,s)$ of Definition 3.1 is a two-parameter process with $(t,s)\in[0,T+\delta]\times[0,T+\delta]$ and its value is separately determined on different time domains. For $(t,s)\in\Delta_\delta^c$, $Z(t,s)$ is defined by (3.4). For $(t,s)\in\Delta$, $Z(t,s)$ is endogenously determined together with Y_t such that (3.1) holds in Itô's sense. Since f depends on $Z(t,s)$ on Δ without anticipation, the equality of (3.1) is independent of the value of Z on $\Delta_\delta\setminus\Delta$, i.e., any value of Z on $\Delta_\delta\setminus\Delta$ equalizes VNBSFE (3.1). For the uniqueness of solution, we define $Z(t,s)=0$ in Definition 3.1 on $\Delta_\delta\setminus\Delta$.

For all $\tau\in[0,T+\delta]$, define a subspace of $\mathscr{H}^2(0,\tau)$,

$$\mathscr{M}^2(0,\tau):=\left\{(\theta,\vartheta)\in\mathscr{H}^2(0,\tau)\;\middle|\;\theta(t)=E[\theta(t)]+\int_0^t\vartheta(t,s)dW(s),\forall\, t\in[0,\tau]\right\}$$

equipped with norm

$$\|(\theta,\vartheta)\|^2_{\mathscr{M}^2(0,\tau)}=E\left[\int_0^\tau|\theta(u)|^2du+\int_0^\tau\int_s^\tau|\vartheta(s,u)|^2duds\right].$$

Then $\mathscr{M}^2(0,\tau)$ is a closed subspace of $\mathscr{H}^2(0,\tau)$ under the norm $\|\cdot\|_{\mathscr{H}^2(0,\tau)}$. In fact, it is also a complete space under $\|\cdot\|_{\mathscr{M}^2(0,\tau)}$, because $\|\cdot\|_{\mathscr{H}^2(0,\tau)}$ is equivalent to $\|\cdot\|_{\mathscr{M}^2(0,\tau)}$ in $\mathscr{M}^2(0,\tau)$. For all $(\theta,\vartheta)\in\mathscr{M}^2(0,\tau)$,

$$E\left[\int_0^t|\vartheta(t,s)|^2ds\right]=E[|\theta(t)-E[\theta(t)]|^2]\leqslant 2E[|\theta(t)|^2].$$

Thus,

$$\begin{aligned}\|(\theta,\vartheta)\|^2_{\mathscr{H}^2(0,\tau)}&=E\left[\int_0^\tau|\theta(t)|^2dt+\int_0^\tau\int_0^\tau|\vartheta(t,s)|^2dsdt\right]\\&\leqslant 2E\left[\int_0^\tau|\theta(t)|^2dt+\int_0^\tau\int_t^\tau|\vartheta(t,s)|^2dsdt\right]=2\|(\theta,\vartheta)\|^2_{\mathscr{M}^2(0,\tau)}\\&\leqslant 2E\left[\int_0^\tau|\theta(t)|^2dt+\int_0^\tau\int_0^\tau|\vartheta(t,s)|^2dsdt\right]=2\|(\theta,\vartheta)\|^2_{\mathscr{H}^2(0,\tau)}.\end{aligned}$$

Therefore, if $(Y,Z)\in\mathscr{H}^2(0,T+\delta)$ is the M-solution to VNBSFE (3.1), it means $(Y,Z)\in\mathscr{M}^2(0,T+\delta)$ and $Z(t,s)=0$ on $\Delta_\delta\setminus\Delta$.

Before showing the existence and uniqueness of M-solution to VNBSFE (3.1), we discuss some backward equations. First, consider the following trivial backward SDE (BSDE),

$$Y(t)=\zeta+\int_t^T f(s)ds-\int_t^T Z(s)dW(s),\quad t\in[0,T], \tag{3.5}$$

where $f\in\mathscr{L}^2_{\mathbb{F}}(0,T;\mathbb{R}^n)$ and $\zeta\in\mathbb{L}^2(\Omega;\mathbb{R}^n)$. The proof of the following Lemma 3.1 is standard (see [10,28]).

Lemma 3.1. *BSDE* (3.5) *admits a unique pair of solution* $(Y,Z)\in\mathscr{S}^2_{\mathbb{F}}([0,T];\mathbb{R}^n)\times\mathscr{L}^2_{\mathbb{F}}(0,T;\mathbb{R}^{n\times d})$, *and for any* $t\in[0,T]$,

$$\mathrm{e}^{\beta t}|Y(t)|^2+E\left[\int_t^T\mathrm{e}^{\beta s}|Z(s)|^2ds\;\middle|\;\mathscr{F}_t\right]\leqslant E\left[2\mathrm{e}^{\beta T}|\zeta|^2+\alpha\int_t^T\mathrm{e}^{\beta s}|f(s)|^2ds\;\middle|\;\mathscr{F}_t\right], \tag{3.6}$$

where α *and* β *are any two positive constants satisfying* $\beta\geqslant\frac{2}{\alpha}$.

Consider the following backward integral equation,

$$\rho(t,s)=\Phi(t)+\int_s^T h(t,u)du-\int_s^T \nu(t,u)dW_u, \quad t\in[0,T], \tag{3.7}$$

where $\Phi\in\mathbb{L}^2(0,T;\mathbb{L}^2(\Omega;\mathbb{R}^n))$, and $h\in\mathbb{L}^2(0,T;\mathscr{L}^2_{\mathbb{F}}(0,T;\mathbb{R}^n))$. Fixed $t\in[0,T]$, (3.7) is a BSDE with generator $h(t,\cdot)\in\mathscr{L}^2_{\mathbb{F}}(0,T;\mathbb{R}^n)$ and terminal condition $\Phi(t)\in\mathbb{L}^2(\Omega;\mathbb{R}^n)$. Thus (3.7) is a family of BSDEs parameterized by $t\in[0,T]$. Let $s=t$. Denote $y(t):=\rho(t,t)$ and $z(t,u):=\nu(t,u)$ when $u\geqslant t$. Then

$$y(t)=\Phi(t)+\int_t^T h(t,u)du-\int_t^T z(t,u)dW_u, \quad t\in[0,T]. \tag{3.8}$$

It is not a BSDE, but a backward stochastic Volterra integral equation (BSVIE), which was studied in [24,35].

Remark 3.2. In (3.8), the equality is independent of z on Δ^c. Then any value of z on Δ^c equalizes (3.8), such as $z(t,u)=\nu(t,u)$ or $z(t,u)=0$, $(t,u)\in\Delta^c$. Therefore, the uniqueness of solution does not hold. However, in the definition of M-solution, the value of z on Δ^c is settled by $y(t)=E[y(t)]+\int_0^t z(t,s)dW(s)$. This determines the uniqueness.

Directly from [35, Theorem 3.7] and estimate (3.6) of Lemma 3.1, we conclude the following lemma.

Lemma 3.2. *BSVIE* (3.8) *admits a unique pair of solution* $(y(\cdot),z(\cdot,\cdot))\in\mathscr{M}^2(0,T)$. *In addition,*

$$E\bigg[e^{\beta t}|y(t)|^2+\int_t^T e^{\beta s}|z(t,s)|^2ds\bigg]\leqslant E\bigg[e^{\beta T}|\Phi(t)|^2+\alpha\int_t^T e^{\beta s}|h(t,s)|^2ds\bigg], \tag{3.9}$$

where $\alpha>0$ *and* $\beta\geqslant\frac{2}{\alpha}$ *are any two positive constants.*

Next theorem is devoted to the existence and uniqueness of the M-solution to VNBSFE (3.1).

Theorem 3.3. *Suppose that* (G,f) *satisfies* (H1)–(H3). *Then for any* $\Psi(\cdot)\in\mathbb{L}^2(0,T;\mathbb{L}^2(\Omega;\mathbb{R}^n))$ *and* $\xi(\cdot)\in\mathscr{L}^2_{\mathbb{F}}(T,T+\delta;\mathbb{R}^n)$, *VNBSFE* (3.1) *admits a unique pair of* M*-solution* $(Y,Z)\in\mathscr{M}^2(0,T+\delta)$. *Moreover, the following estimate holds*:

$$\begin{aligned}&E\bigg[\int_0^{T+\delta}|Y(t)|^2dt+\int_0^{T+\delta}\int_t^{T+\delta}|Z(t,s)|^2dsdt\bigg]\\&\leqslant CE\bigg[\int_0^T|\Psi(t)|^2dt+\int_T^{T+\delta}|\xi(t)|^2dt+\int_0^T|G(t,0)|^2dt+\int_0^T\int_t^T|f_0(t,s)|^2dsdt\bigg].\end{aligned} \tag{3.10}$$

Proof. **Step 1.** Define a subset of $\mathscr{M}^2(0,T+\delta)$,

$$\mathscr{M}^2_\xi(0,T):=\{(\theta,\vartheta)\in\mathscr{M}^2(0,T+\delta)\mid\theta(t)=\xi(t),\forall\, t\in(T,T+\delta],\text{and }\vartheta(t,s)=0,\forall\,(t,s)\in\Delta_\delta\setminus\Delta\}$$

equipped with the norm

$$\|(\theta,\vartheta)\|^2=E\bigg[\int_0^T e^{\beta t}|\theta(t)|^2dt+\int_0^T\int_t^T e^{\beta s}|\vartheta(t,s)|^2dsdt\bigg],$$

where β is a positive constant waiting to be determined later. It is obvious that $\mathscr{M}^2_\xi(0,T)$ is a closed subset of $\mathscr{M}^2(0,T+\delta)$.

For each $(y(\cdot),z(\cdot,\cdot))\in\mathscr{M}^2_\xi(0,T)$, consider

$$\begin{cases}Y(t)-G(t,y_t)=\Psi(t)+\int_t^T f(t,s,y_s,z(t,s),z(s,t;\delta))ds-\int_t^T Z(t,s)dW_s, & t\in[0,T],\\ Y(t)=\xi(t), & t\in(T,T+\delta].\end{cases} \tag{3.11}$$

Denote $\tilde{Y}(t):=Y(t)-G(t,y_t)$, then

$$\tilde{Y}(t)=\Psi(t)+\int_t^T f(t,s,y_s,z(t,s),z(s,t;\delta))ds-\int_t^T Z(t,s)dW_s, \quad t\in[0,T]. \tag{3.12}$$

Since $(y,z)\in\mathscr{M}^2_\xi(0,T)$, $f(t,s,y_s,z(t,s),z(s,t;\delta))\in\mathbb{L}^2(0,T;\mathscr{L}^2_{\mathbb{F}}(0,T;\mathbb{R}^{n\times m}))$ due to (H1) and (H2).

By Lemma 3.2, (3.12) admits a unique pair of solution $(\tilde{Y},Z)\in\mathscr{M}^2(0,T)$. Define

$$Y(t)=\begin{cases}\tilde{Y}(t)-G(t,y_t), & t\in[0,T],\\ \xi(t), & t\in(T,T+\delta].\end{cases}$$

Then $Y\in\mathscr{L}^2_{\mathbb{F}}(0,T+\delta;\mathbb{R}^n)$. Define $Z(t,s)=0$ on $\Delta_\delta\setminus\Delta$ and modify the value of Z on Δ_δ such that

$$Y(t)=E[Y(t)]+\int_0^t Z(t,s)dW_s,\quad \forall\, t\in[0,T+\delta].$$

Then $(Y,Z)\in\mathscr{M}^2_\xi(0,T)$ is an M-solution to (3.11).

Step 2. Consider the mapping $\Gamma:(y(\cdot),z(\cdot,\cdot))\mapsto(Y(\cdot),Z(\cdot,\cdot))$ with (Y,Z) being defined associated with (y,z) in Step 1. We shall prove that Γ is a contraction mapping on Banach space $\mathscr{M}^2_\xi(0,T)$.

Take another pair $(\bar{y}(\cdot),\bar{z}(\cdot,\cdot))\in\mathscr{M}^2_\xi(0,T)$, and denote $(\bar{Y}(\cdot),\bar{Z}(\cdot,\cdot))\in\mathscr{M}^2_\xi(0,T)$ as the M-solution to (3.11) with $(y(\cdot),z(\cdot))$ replaced by $(\bar{y}(\cdot),\bar{z}(\cdot))$. Define $\Delta Y(t):=Y(t)-\bar{Y}(t)$, $\Delta Z(t,s):=Z(t,s)-\bar{Z}(t,s)$, $\Delta y(t):=y(t)-\bar{y}(t)$ and $\Delta z(t):=z(t,s)-\bar{z}(t,s)$. Then

$$\begin{aligned}&\Delta Y(t)-[G(t,y_t)-G(t,\bar{y}_t)]\\&\quad=\int_t^T[f(t,s,y_s,z(t,s),z(s,t;\delta))-f(t,s,\bar{y}_s,\bar{z}(t,s),\bar{z}(s,t;\delta))]ds+\int_t^T\Delta Z(t,s)dW(s).\end{aligned}$$

We shall denote by C a constant which may vary from line to line. In view of (3.9) in Lemma 3.2 and choosing $\beta\geqslant\frac{2}{\alpha}$, we have

$$\begin{aligned}&E\bigg[\mathrm{e}^{\beta t}|\Delta Y(t)-[G(t,y_t)-G(t,\bar{y}_t)]|^2+\int_t^T\mathrm{e}^{\beta s}|\Delta Z(t,s)|^2ds\bigg]\\&\quad\leqslant\alpha E\bigg[\int_t^T\mathrm{e}^{\beta s}|f(t,s,y_s,z(t,s),z(s,t;\delta))-f(t,s,\bar{y}_s,\bar{z}(t,s),\bar{z}(s,t;\delta))|^2ds\bigg]\\&\quad\leqslant\alpha CE\bigg\{\int_t^T\mathrm{e}^{\beta s}\bigg[\int_0^\delta|\Delta y(s+u)|^2\varrho_1(du)+|\Delta z(t,s)|^2+\int_0^\delta|\Delta z(s+u,t+u)|^2\varrho_2(du)\bigg]ds\bigg\}.\end{aligned}\tag{3.13}$$

Integrating (3.13) in t from 0 to T, and denoting $\Delta G(t):=G(t,y_t)-G(t,\bar{y}_t)$, we have

$$\begin{aligned}&E\bigg[\int_0^T\mathrm{e}^{\beta t}|\Delta Y(t)-\Delta G(t)|^2dt+\int_0^T\int_t^T\mathrm{e}^{\beta s}|\Delta Z(t,s)|^2dsdt\bigg]\\&\quad\leqslant\alpha CE\bigg\{\int_0^T\int_t^T\mathrm{e}^{\beta s}\bigg[\int_0^\delta|\Delta y(s+u)|^2\varrho_1(du)+|\Delta z(t,s)|^2+\int_0^\delta|\Delta z(s+u,t+u)|^2\varrho_2(du)\bigg]dsdt\bigg\}\\&\quad\leqslant\alpha CE\bigg[\int_0^T\mathrm{e}^{\beta s}|\Delta y(s)|^2ds+\int_0^T\int_t^T\mathrm{e}^{\beta s}(|\Delta z(t,s)|^2+|\Delta z(s,t)|^2)dsdt\bigg]\\&\quad\leqslant\alpha C\,E\bigg[\int_0^T\mathrm{e}^{\beta s}|\Delta y(s)|^2ds+\int_0^T\int_t^T\mathrm{e}^{\beta s}|\Delta z(t,s)|^2dsdt\bigg].\end{aligned}\tag{3.14}$$

The last inequality is due to the fact

$$E\bigg[\int_0^T\int_t^T\mathrm{e}^{\beta s}|\Delta z(s,t)|^2dsdt\bigg]=E\bigg[\int_0^T\mathrm{e}^{\beta t}\int_0^t|\Delta z(t,s)|^2dsdt\bigg]\leqslant E\bigg[\int_0^T\mathrm{e}^{\beta t}|\Delta y(t)|^2dt\bigg].$$

Since for all $\gamma\in(0,1)$ and $a,b\in\mathbb{R}^n$, $|a-b|^2\geqslant(1-\gamma)|a|^2-(\frac{1}{\gamma}-1)|b|^2$, then

$$|\Delta Y(t)-\Delta G(t)|^2\geqslant(1-\gamma)|\Delta Y(t)|^2-\bigg(\frac{1}{\gamma}-1\bigg)|\Delta G(t)|^2.$$

(3.14) is reduced to

$$
\begin{aligned}
&E\bigg[(1-\gamma)\int_0^T \mathrm{e}^{\beta t}|\Delta Y(t)|^2dt+\int_0^T\int_t^T \mathrm{e}^{\beta s}|\Delta Z(t,s)|^2dsdt\bigg]\\
&\quad\leqslant\left(\frac{1}{\gamma}-1\right)E\bigg[\int_0^T\mathrm{e}^{\beta t}|\Delta G(t)|^2dt\bigg]+\alpha C\,E\bigg[\int_0^T\mathrm{e}^{\beta t}|\Delta y(t)|^2dt+\int_0^T\int_t^T\mathrm{e}^{\beta s}|\Delta z(t,s)|^2dsdt\bigg]\\
&\quad\leqslant\bigg[\left(\frac{1}{\gamma}-1\right)\kappa+\alpha C\bigg]E\bigg[\int_0^T\mathrm{e}^{\beta t}|\Delta y(t)|^2ds\bigg]+\alpha C\,E\bigg[\int_0^T\int_t^T\mathrm{e}^{\beta s}|\Delta z(t,s)|^2dsdt\bigg].
\end{aligned}
$$

To prove Γ is a contraction mapping, it suffices to show: For all $\kappa\in(0,1)$, there is $\gamma\in(0,1)$ such that

$$
\left(\frac{1}{\gamma}-1\right)\kappa+\alpha C<1-\gamma \quad \text{and} \quad \alpha C<1,
$$

which hold true by choosing α to be small sufficiently. Therefore, Γ admits a unique fixed point (Y,Z) $\in\mathscr{M}_\xi^2(0,T)$, which proves to be the unique M-solution to VNBSFE (3.1).

Step 3. Similar to the proof in Step 2, we have for all $\alpha\in(0,1)$ and $M>0$,

$$
\begin{aligned}
&E\bigg[(1-\gamma)\int_0^T \mathrm{e}^{\beta t}|Y(t)|^2dt+\int_0^T\int_t^T \mathrm{e}^{\beta s}|Z(t,s)|^2dsdt\bigg]\\
&\quad\leqslant E\bigg[\mathrm{e}^{\beta T}\int_0^T|\Psi(t)|^2dt+\left(\frac{1}{\gamma}-1\right)\int_0^T\mathrm{e}^{\beta t}|G(t,Y_t)|^2dt+\alpha C\int_0^T\int_t^T\mathrm{e}^{\beta s}|f_0(t,s)|^2dsdt\bigg]\\
&\qquad+\alpha C\,E\bigg[\int_0^{T+\delta}\mathrm{e}^{\beta t}|Y(t)|^2dt+\int_0^T\int_t^T\mathrm{e}^{\beta s}|Z(t,s)|^2dsdt\bigg]\\
&\quad\leqslant E\bigg[\mathrm{e}^{\beta T}\int_0^T|\Psi(t)|^2dt+\left(\frac{1}{\gamma}-1\right)(1+M)\int_0^T\mathrm{e}^{\beta t}|G(t,0)|^2dt+\alpha C\int_0^T\int_t^T\mathrm{e}^{\beta s}|f_0(t,s)|^2dsdt\\
&\qquad+\bigg[\left(\frac{1}{\gamma}-1\right)\left(1+\frac{1}{M}\right)\kappa+\alpha C\bigg]\int_0^T\mathrm{e}^{\beta t}|Y(t)|^2dt+\alpha C\int_0^T\int_t^T\mathrm{e}^{\beta s}|Z(t,s)|^2dsdt\bigg].
\end{aligned}
$$

It is easy to prove that there are $\gamma\in(0,1)$ and $M>0$, such that the following two inequalities hold for any $\kappa\in(0,1)$ by choosing α small sufficiently,

$$
\left(\frac{1}{\gamma}-1\right)\left(1+\frac{1}{M}\right)\kappa+\alpha C<1-\gamma \quad \text{and} \quad \alpha C<1.
$$

Then the estimate (3.10) holds. □

4 Dual analysis

In this section, we are concerned with the dual analysis between linear NSFDEs and VNBSFEs which is crucial in the Pontryagin maximum principle. To comprehend further the roles of the duality, the readers can skip to the next section. In fact, the dual analysis herein can be extended to the infinite dimensional cases without any essential difficulties.

Denote $(\bar{X}(\cdot),\bar{u}(\cdot))\in\mathscr{S}_{\mathbb{F}}^2([-\delta,T];\mathbb{R}^n)\times\mathcal{U}_{ad}$ as an optimal pair, and $v\in\mathscr{L}_{\mathbb{F}}^2(0,T;\mathbb{R}^m)$. Let $\chi(\cdot)$ be the solution to the following linear NSFDE:

$$
\begin{cases}
d[\chi(t)-\bar{g}_x(t)\chi^t]=[\bar{b}_x(t)\chi^t+\bar{b}_u(t)v(t)]dt+[\bar{\sigma}_x(t)\chi^t+\bar{\sigma}_u(t)v(t)]dW(t), \quad t\in[0,T],\\
\chi(t)=0, \quad t\in[-\delta,0].
\end{cases}
\tag{4.1}
$$

We note that, linear NSFDE of the form (4.1) comes from the first order variation of the controlled NSFDE (1.2) (see also Lemma 5.1 below). Consider the linear functional

$$
I(\chi(\cdot))=E\int_0^T\bar{l}_x(t)\chi^tdt.
$$

Here, $\bar{g}_x(t) := g_x(t, \bar{X}^t)$, $(\bar{b}_x, \bar{b}_u, \bar{\sigma}_x, \bar{\sigma}_u, \bar{l}_x)(t) := (b_x, b_u, \sigma_x, \sigma_u, l_x)(t, \bar{X}^t, \bar{u}(t))$.

Under Assumptions (A1)–(A3), (4.1) is well-posed and we have

$$\chi(t) - \int_{-\delta}^{0} \bar{G}(t,r)\chi(t+r)\lambda_0(dr) = \int_0^t \left[\int_{-\delta}^{0} \bar{B}(s,r)\chi(s+r)\lambda_1(dr) + \bar{b}_u(s)v(s)\right]ds + \int_0^t \left[\int_{-\delta}^{0} \bar{\varsigma}(s,r)\chi(s+r)\lambda_2(dr) + \bar{\sigma}_u(s)v(s)\right]dW(s), \tag{4.2}$$

and

$$I(\chi(\cdot)) = E\int_0^T \int_{-\delta}^{0} \bar{L}(t,r)\chi(t+r)\lambda_3(dr)dt,$$

where, for simplicity, we set

$$(\bar{B}, \bar{\varsigma}, \bar{L})(t,r) := (\bar{B}, \bar{\varsigma}, \bar{L})(t, \bar{X}^t, \bar{u}(t), r) \quad \text{and} \quad \bar{G}(t,r) := \bar{G}(t, \bar{X}^t, r).$$

Denote

$$\rho(t) := \int_0^t \bar{b}_u(s)v(s)ds + \int_0^t \bar{\sigma}_u(s)v(s)dW(s).$$

We have the following duality:

Proposition 4.1. *Let $\chi \in \mathscr{S}^2_{\mathbb{F}}([-\delta, T]; \mathbb{R}^n)$ be the solution to NSFDE (4.2), and $(Y, Z) \in \mathscr{M}^2(0, T+\delta)$ be the M-solution to the following linear VNBSFE:*

$$\begin{cases} Y(t) - E_t\left[\int_{-\delta}^{0} \bar{G}'(t-r,r)Y(t-r)\lambda_0(dr)\right] \\ \qquad = \int_{-\delta}^{0} \bar{L}(t-r,r)\lambda_3(dr) + \int_t^T E_s\Big[\int_{-\delta}^{0} \bar{B}'(t-r,r)Y(s-r)\lambda_1(dr) \\ \qquad\quad + \int_{-\delta}^{0} \langle \bar{\varsigma}(t-r,r),\, Z(s-r,t-r)\rangle \lambda_2(dr)\Big]ds - \int_t^T Z(t,s)dW(s), \quad t \in [0,T], \\ Y(t) = 0, \quad t \in (T, T+\delta], \end{cases} \tag{4.3}$$

where we set $(\bar{G}, \bar{B}, \bar{\varsigma}, \bar{L})(t, \cdot) = 0$, $\forall\, t \in (T, T+\delta]$. Then the following relation holds:

$$I(\chi(\cdot)) = E\left[\int_0^T \bar{l}_x(t)\chi^t dt\right] = E\left[\int_0^T \langle \rho(t),\, Y(t)\rangle dt\right].$$

Note that the well-posedness of VNBSFE (4.3) is ensured by Theorem 3.3.

Proof. In view of (4.2), we have

$$\rho(t) = \chi(t) - \int_{-\delta}^{0} \bar{G}(t,r)\chi(t+r)\lambda_0(dr) - \int_0^t\int_{-\delta}^{0} \bar{B}(s,r)\chi(s+r)\lambda_1(dr)ds - \int_0^t\int_{-\delta}^{0} \bar{\varsigma}(s,r)\chi(s+r)\lambda_2(dr)dW(s).$$

Since $Y \in \mathscr{L}^2_{\mathbb{F}}(0, T; \mathbb{R}^n)$, we get

$$\begin{aligned} E\left[\int_0^T \langle Y(t), \rho(t)\rangle dt\right] &= E\left[\int_0^T \langle Y(t), \chi(t)\rangle dt\right] - E\left[\int_0^T \left\langle Y(t),\, \int_{-\delta}^{0} \bar{G}(t,r)\chi(t+r)\lambda_0(dr)\right\rangle dt\right] \\ &\quad - E\left[\int_0^T \left\langle Y(t),\, \int_0^t\int_{-\delta}^{0} \bar{B}(s,r)\chi(s+r)\lambda_1(dr)ds\right\rangle dt\right] \\ &\quad - E\left[\int_0^T \left\langle Y(t),\, \int_0^t\int_{-\delta}^{0} \bar{\varsigma}(s,r)\chi(s+r)\lambda_2(dr)dW(s)\right\rangle dt\right] \\ &= E\left[\int_0^T \langle Y(t), \chi(t)\rangle dt\right] - I_1 - I_2 - I_3. \end{aligned}$$

By Fubini's theorem, we have

$$
\begin{aligned}
I_1 &= \int_{-\delta}^{0}\int_{0}^{T} \langle Y(t),\, \bar{G}(t,r)\chi(t+r)\rangle dt\, \lambda_0(dr) \\
&= \int_{-\delta}^{0}\int_{r}^{T+r} \langle \bar{G}'(t-r,r)Y(t-r),\, \chi(t)\rangle dt\, \lambda_0(dr) \\
&\quad (\text{Noting that } \chi(t)1_{[-\delta,0]}(t)=0 \text{ and } Y(t)1_{(T,T+\delta]}(t)=0) \\
&= \int_{-\delta}^{0}\int_{0}^{T} \langle \bar{G}'(t-r,r)Y(t-r), \chi(t)\rangle dt\, \lambda_0(dr) \\
&= \int_{0}^{T} \left\langle \int_{-\delta}^{0} \bar{G}'(t-r,r)Y(t-r)\lambda_0(dr), \chi(t) \right\rangle dt
\end{aligned} \tag{4.4}
$$

and

$$
\begin{aligned}
I_2 &= \int_{-\delta}^{0}\int_{0}^{T} \left\langle \bar{B}'(t,r)\int_{t}^{T} Y(s)ds,\, \chi(t+r) \right\rangle dt\, \lambda_1(dr) \\
&= \int_{-\delta}^{0}\int_{r}^{T+r} \left\langle \bar{B}'(t-r,r)\int_{t}^{T} Y(s-r)ds,\, \chi(t) \right\rangle dt\, \lambda_1(dr) \\
&= \int_{0}^{T} \left\langle \int_{t}^{T}\int_{-\delta}^{0} \bar{B}'(t-r,r)Y(s-r)\lambda_1(dr)ds,\, \chi(t) \right\rangle dt.
\end{aligned} \tag{4.5}
$$

In view of $(Y,Z)\in \mathscr{M}^2(0,T+\delta)$, and $Y(t)=E[Y(t)]+\int_0^t Z(t,s)dW(s)$, we further have

$$
\begin{aligned}
I_3 &= \int_{0}^{T} \left\langle \int_{0}^{t} Z(t,s)dW(s),\, \int_{0}^{t}\int_{-\delta}^{0} \bar{\varsigma}(s,r)\chi(s+r)\lambda_2(dr)dW(s) \right\rangle dt \\
&= \int_{0}^{T}\int_{0}^{t} \left\langle Z(t,s),\, \int_{-\delta}^{0} \bar{\varsigma}(s,r)\chi(s+r)\lambda_2(dr) \right\rangle ds dt \\
&= \int_{-\delta}^{0}\int_{0}^{T} \left\langle \int_{t}^{T} \langle \bar{\varsigma}(t,r),\, Z(s,t)\rangle ds,\, \chi(t+r) \right\rangle dt\, \lambda_2(dr) \\
&= \int_{-\delta}^{0}\int_{r}^{T+r} \left\langle \int_{t}^{T} \langle \bar{\varsigma}(t-r,r),\, Z(s-r,t-r)\rangle ds,\, \chi(t) \right\rangle dt\, \lambda_2(dr) \\
&= \int_{0}^{T} \left\langle \int_{t}^{T}\int_{-\delta}^{0} \langle \bar{\varsigma}(t-r,r),\, Z(s-r,t-r)\rangle \lambda_2(dr)ds,\, \chi(t) \right\rangle dt.
\end{aligned} \tag{4.6}
$$

Combining (4.4)–(4.6), we obtain

$$
\begin{aligned}
E\left[\int_0^T \langle Y(t),\rho(t)\rangle dt\right] &= E\left[\int_0^T \left\langle \int_{-\delta}^{0} \bar{L}(t-r,r)\lambda_3(dr) - \int_t^T Z(t,s)dW(s),\, \chi(t) \right\rangle dt\right] \\
&= E\left[\int_0^T \left\langle \int_{-\delta}^{0} \bar{L}(t-r,r)\lambda_3(dr),\, \chi(t) \right\rangle dt\right] \\
&= E\left[\int_{-\delta}^{0}\int_{-r}^{T-r} \langle \bar{L}(t,r),\, \chi(t+r)\rangle dt\, \lambda_3(dr)\right] \\
&= E\left[\int_0^T \int_{-\delta}^{0} \bar{L}(t,r)\chi(t+r)\, \lambda_3(dr)dt\right] = I(\chi(\cdot)). \qquad \square
\end{aligned}
$$

5 Maximum principle

In this section, we shall establish the Pontryagin maximum principle for the optimal control problem in Section 2.

Suppose that $(\bar{X}(\cdot), \bar{u}(\cdot))$ is an optimal pair. For any $u(\cdot) \in \mathcal{U}_{ad}$, denote $v(\cdot) := u(\cdot) - \bar{u}(\cdot)$ and

$$u_\varepsilon(\cdot) := \bar{u}(\cdot) + \varepsilon v(\cdot) \in \mathcal{U}_{ad}, \quad \forall\, \varepsilon \in [0,1].$$

Denote by $X_\varepsilon(\cdot)$ the solution to NSFDE (2.1) associated with control u_ε.

Before constructing the maximum principle, we need the following estimate of the first order expansion.

Lemma 5.1. *Suppose that assumptions* (A1) *and* (A2) *hold. Then we have the following first order expansion*:

$$X_\varepsilon(t) = \bar{X}(t) + \varepsilon\chi(t) + R_\varepsilon(t), \quad t \in [0,T]$$

with

$$\lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon^2} E\Big[\sup_{0\leqslant t\leqslant T} |R_\varepsilon(t)|^2\Big] = 0.$$

Here, χ satisfies the following linear NSFDE:

$$\begin{cases} d[\chi(t) - \bar{g}_x(t)\chi^t] = [\bar{b}_x(t)\chi^t + \bar{b}_u(t)v(t)]dt + [\bar{\sigma}_x(t)\chi^t + \bar{\sigma}_u(t)v(t)]dW(t), \quad t \in [0,T], \\ \chi(t) = 0, \quad t \in [-\delta, 0] \end{cases} \tag{5.1}$$

with $\bar{g}_x(t) := g_x(t, \bar{X}^t)$ *and* $(\bar{b}_x, \bar{b}_u, \bar{\sigma}_x, \bar{\sigma}_u)(t) := (b_x, b_u, \sigma_x, \sigma_u)(t, \bar{X}^t, \bar{u}(t))$.

Proof. Setting $z_\varepsilon(t) := \frac{X_\varepsilon(t) - \bar{X}(t)}{\varepsilon}$, $\forall\, t \in [-\delta, T]$, we have $R_\varepsilon(t) = \varepsilon(z_\varepsilon - \chi)(t)$ and

$$\begin{aligned} &z_\varepsilon(t) - \int_0^1 g_x(t, \bar{X}^s + \theta\varepsilon z_\varepsilon^t)\, z_\varepsilon^t d\theta \\ &\quad= \int_0^t \Big[\int_0^1 b_x(s, \bar{X}^s + \theta\varepsilon z_\varepsilon^t, u_\varepsilon(s)) z_\varepsilon^s d\theta + \int_0^1 b_u(s, \bar{X}^s, \bar{u}(s) + \varepsilon\theta\bar{v}(s))\, \bar{v}(s) d\theta\Big] ds \\ &\qquad+ \int_0^t \Big[\int_0^1 \sigma_x(s, \bar{X}^s + \theta\varepsilon z_\varepsilon^t, u_\varepsilon(s))\, z_\varepsilon^s d\theta + \int_0^1 \sigma_u(s, \bar{X}^s, \bar{u}(s) + \varepsilon\theta\bar{v}(s))\, \bar{v}(s) d\theta\Big] dW(s). \end{aligned}$$

Thus, by Assumptions (A1) and (A2),

$$\begin{aligned} &(1-\kappa) E \sup_{s\in[-\delta,t]} |z_\varepsilon(s)|^2 \\ &\quad\leqslant E\Big[\sup_{s\in[-\delta,t]} \Big| z_\varepsilon(t) - \int_0^1 g_x(t, \bar{X}^s + \theta\varepsilon z_\varepsilon^t)\, z_\varepsilon^t d\theta\Big|^2\Big] \\ &\quad= E\Big[\sup_{s\in[-\delta,t]} \Big| \int_0^t \Big[\int_0^1 b_x(s, \bar{X}^s + \theta\varepsilon z_\varepsilon^t, u_\varepsilon(s)) z_\varepsilon^s d\theta + \int_0^1 b_u(s, \bar{X}^s, \bar{u}(s) + \varepsilon\theta\bar{v}(s))\, \bar{v}(s) d\theta\Big] ds \\ &\qquad+ \int_0^t \Big[\int_0^1 \sigma_x(s, \bar{X}^s + \theta\varepsilon z_\varepsilon^t, u_\varepsilon(s))\, z_\varepsilon^s d\theta + \int_0^1 \sigma_u(s, \bar{X}^s, \bar{u}(s) + \varepsilon\theta\bar{v}(s))\, \bar{v}(s) d\theta\Big] dW(s)\Big|^2\Big] \\ &\quad\leqslant C\, E\Big[\int_0^t \sup_{r\in[-\delta,s]} |z_\varepsilon(r)|^2 ds + \int_0^t |\bar{v}(s)|^2 ds\Big], \end{aligned} \tag{5.2}$$

which, by Gronwall's inequality, implies

$$E \sup_{s\in[-\delta,T]} |z_\varepsilon(s)|^2 \leqslant CE \int_0^T |\bar{v}(s)|^2 ds. \tag{5.3}$$

Define

$$y_\varepsilon = \frac{R_\varepsilon}{\varepsilon}, \quad g_x^\varepsilon(s) = \int_0^1 g_x(t, \bar{X}^s + \theta\varepsilon z_\varepsilon^t) d\theta - g_x(t, \bar{X}^s)$$

and

$$(b_x^\varepsilon, \sigma_x^\varepsilon)(s) = \int_0^1 (b_x, \sigma_x)(s, \bar{X}^s + \theta\varepsilon z_\varepsilon^t, u_\varepsilon(s)) d\theta - (\bar{b}_x, \bar{\sigma}_x)(s),$$

$$(b_u^\varepsilon,\sigma_u^\varepsilon)(s)=\int_0^1(b_u,\sigma_u)(s,\bar X^s,\bar u(s)+\varepsilon\theta\bar v(s))d\theta-(\bar b_u,\bar\sigma_u)(s).$$

It follows that

$$\begin{aligned}&y_\varepsilon(t)-\bar g_x\,y_\varepsilon^t-g_x^\varepsilon(t)z_\varepsilon^t\\&\quad=\int_0^t[\bar b_xy_\varepsilon^s+b_x^\varepsilon(s)z_\varepsilon^s+b_u^\varepsilon(s)\bar v(s)]ds+\int_0^t[\bar\sigma_xy_\varepsilon^s+\sigma_x^\varepsilon(s)z_\varepsilon^s+\sigma_u^\varepsilon(s)\bar v(s)]dW(s).\end{aligned}$$

In a similar way to (5.2) and (5.3), we obtain

$$\begin{aligned}&E\sup_{s\in[-\delta,T]}|y_\varepsilon(s)|^2\\&\leqslant CE\bigg[\int_0^T(|b_x^\varepsilon(s)z_\varepsilon^s|^2+|\sigma_x^\varepsilon(s)z_\varepsilon^s|^2+|b_u^\varepsilon(s)\bar v(s)|^2+|\sigma_u^\varepsilon(s)\bar v(s)|^2)ds+\sup_{s\in[0,T]}|g_x^\varepsilon(s)z_\varepsilon^s|^2\bigg]\\&\to0,\quad\text{ as }\varepsilon\to0^+\quad\text{(by Lebesgue domination convergence theorem).}\end{aligned}$$

Hence,

$$\lim_{\varepsilon\to0^+}\frac{1}{\varepsilon^2}E\Big[\sup_{0\leqslant t\leqslant T}|R_\varepsilon(t)|^2\Big]=0.$$

We complete the proof. □

Now, it is the stage to establish the Pontryagin maximum principle. Define the Hamiltonian function

$$H(t,\psi,u;P,Q)=l(t,\psi,u)+b'(t,\psi,u)E_t\bigg[\int_t^TP(s)ds\bigg]+\bigg\langle\sigma(t,\psi,u),\int_t^TQ(s,t)ds\bigg\rangle,\tag{5.4}$$

$\psi\in C([-\delta,0],\mathbb{R}^n),(P,Q)\in\mathscr{H}^2(0,T)$. For simplicity, denote

$$(\bar B,\bar\varsigma,\bar L)(t,r):=(\bar B,\bar\varsigma,\bar L)(t,\bar X^t,\bar u(t),r),\quad\bar G(t,r):=\bar G(t,\bar X^t,r),\quad\forall\,(t,r)\in[0,T]\times[-\delta,0],$$

and set $(\bar G,\bar B,\bar\varsigma,\bar L)(t,\cdot)=0$, $\forall\,t\in(T,T+\delta]$. Then we have the following theorem.

Theorem 5.2. *Suppose that assumptions* (A1)–(A3) *hold. Let* $(\bar X(\cdot),\bar u(\cdot))$ *be an optimal pair. Then there exists* $(Y,Z)\in\mathscr{M}^2(0,T+\delta)$ *being the M-solution to the following linear VNBSFE*:

$$\begin{cases}Y(t)-E_t\bigg[\displaystyle\int_{-\delta}^0\bar G'(t-r,r)Y(t-r)\lambda_0(dr)\bigg]\\\qquad=\displaystyle\int_{-\delta}^0\bar L(t-r,r)\lambda_3(dr)+\int_t^TE_s\bigg[\int_{-\delta}^0\bar B'(t-r,r)Y(s-r)\lambda_1(dr)\\\qquad\quad+\displaystyle\int_{-\delta}^0\langle\bar\varsigma(t-r,r),\,Z(s-r,t-r)\rangle\,\lambda_2(dr)\bigg]ds-\int_t^TZ(t,s)dW(s),\quad t\in[0,T],\\Y(t)=0,\quad t\in(T,T+\delta],\end{cases}\tag{5.5}$$

such that the following maximum condition holds:

$$\langle H_u(t,\bar X^t,\bar u(t);Y,Z),u-\bar u(t)\rangle\geqslant0,$$

i.e.,

$$\bigg\langle\bar l_u(t)+\bar b_u'(t)E_t\bigg[\int_t^TY(s)ds\bigg]+\int_t^T\langle\bar\sigma_u(t),\,Z(s,t)\rangle ds,\,u-\bar u(t)\bigg\rangle\geqslant0,\quad dP(\omega)\otimes dt\text{-}a.e.,\quad\forall\,u\in U,$$

where $\bar l_u(t):=l(t,\bar X^t,\bar u(t))$.

Proof. In view of assumptions (A1)–(A3) and Theorem 3.3, VNBSFE (5.5) admits a unique pair of M-solution $(Y,Z)\in\mathscr{M}^2(0,T+\delta)$. Then,

$$\begin{aligned}0&\leqslant\frac{J(u_\varepsilon(\cdot))-J(\bar u(\cdot))}{\varepsilon}=E\int_0^T\frac{l(t,X^t_\varepsilon,u_\varepsilon(t))-l(t,\bar X^t,\bar u(t))}{\varepsilon}dt\\&=E\int_0^T\int_0^1 l_u(t,X^t_\varepsilon,\bar u(t)+\theta\varepsilon v(t))v(t)d\theta dt+E\int_0^T\int_0^1 l_x(t,\bar X^t+\theta(X^t_\varepsilon-\bar X^t),\bar u(t))\frac{X^t_\varepsilon-\bar X^t}{\varepsilon}d\theta dt\\&=E\int_0^T[\bar l_u(t)v(t)+\bar l_x(t)\chi^t]dt+E\int_0^T\int_0^1[l_u(t,X^t_\varepsilon,\bar u(t)+\theta\varepsilon v(t))-\bar l_u(t)]v(t)d\theta dt\\&\quad+E\int_0^T\bigg[\int_0^1 l_x(t,\bar X^t+\theta(X^t_\varepsilon-\bar X^t),\bar u(t))\frac{X^t_\varepsilon-\bar X^t}{\varepsilon}d\theta-\bar l_x(t)\chi^t\bigg]dt,\end{aligned}$$

which, together with Lemma 5.1, implies

$$E\int_0^T[\bar l_u(t)v(t)+\bar l_x(t)\chi^t]dt\geqslant 0.$$

Applying assumption (A3) and Proposition 4.1, we obtain

$$\begin{aligned}0&\leqslant E\int_0^T\bigg[\int_{-\delta}^0\bar L(t,r)\chi(t+r)\lambda_3(dr)+\langle\bar l_u(t),v(t)\rangle\bigg]dt\\&=E\int_0^T[\langle Y(t),\,\rho(t)\rangle+\langle\bar l_u(t),\,v(t)\rangle]dt\\&=E\bigg[\int_0^T\langle\bar l_u(t),\,v(t)\rangle dt+\int_0^T\bigg\langle Y(t),\,\int_0^t\bar b_u(s)v(s)ds\bigg\rangle dt\bigg]\\&\qquad+E\int_0^T\bigg\langle Y(t),\,\int_0^t\bar\sigma_u(s)v(s)dW(s)\bigg\rangle dt\\&=E\int_0^T\bigg\langle\bar l_u(t)+\bar b'_u(t)\int_t^T Y(s)ds+\int_t^T\langle\bar\sigma'_u(t),\,Z(s,t)\rangle ds,\,u(t)-\bar u(t)\bigg\rangle dt.\end{aligned}$$

The last equality is due to

$$\begin{aligned}&E\int_0^T\bigg\langle Y(t),\,\int_0^t\bar\sigma_u(s)v(s)dW(s)\bigg\rangle dt\\&\quad=E\int_0^T\bigg\langle\int_0^t Z(t,s)dW(s),\,\int_0^t\bar\sigma_u(s)v(s)dW(s)\bigg\rangle dt\\&\quad=E\int_0^T\int_0^t\langle Z(t,s),\,\bar\sigma_u(s)v(s)\rangle dsdt=E\int_0^T\bigg\langle\int_t^T\langle\bar\sigma_u(t),\,Z(s,t)\rangle ds,\,v(t)\bigg\rangle dt.\end{aligned}$$

Therefore, for all $u\in U$, there holds

$$\bigg\langle\bar l_u(t)+\bar b'_u(t)E_t\bigg[\int_t^T Y(s)ds\bigg]+\int_t^T\langle\bar\sigma_u(t),\,Z(s,t)\rangle ds,\,u-\bar u(t)\bigg\rangle\geqslant 0,\quad dP(\omega)\otimes dt\text{-a.e..}$$

We complete the proof. □

Remark 5.1. In this work, for simplicity we consider the stochastic optimal control problem (1.1) with the cost functional of Lagrange type and the controlled NSFDE with uncontrolled g. As a matter of fact, without any significant mathematical challenges, we can extend the discussions in Theorem 5.2 to the case with controlled g and general cost functionals, while the arguments will become cumbersome. Indeed, in a similar way to [35, Section 5], we are allowed to extend our results herein without essential difficulties to the optimal control problems with the general cost functional of Bolza type (see [35, (5.2)]). If g depends on u, assume that g is continuously differentiable in u with bounded and continuous derivative, and $\bar g_u(t):=g_u(t,\bar X^t,\bar u(t))$ is continuously in t. Define the admissible control set as follow:

$$\mathcal{U}_{ad}:=\{u:[0,T]\times\Omega\to U,\text{ path-continuous and bounded, }\mathbb{F}\text{-progressively measurable}\},$$

where the path-continuity and boundedness are assumed for simplicity. The maximum principle can be derived similarly, i.e., for all $u \in U$ and almost all $t \in [0, T]$,

$$\left\langle \bar{g}'_u(t)Y(t) + \bar{l}_u(t) + \bar{b}'_u(t)E_t\left[\int_t^T Y(s)ds\right] + \int_t^T \langle \bar{\sigma}_u(t),\, Z(s,t)\rangle ds,\, u - \bar{u}(t) \right\rangle \geqslant 0,$$

and

$$\left\langle \bar{g}'_u(0)E\int_0^T Y(t)dt,\, u(0) - \bar{u}(0) \right\rangle \geqslant 0.$$

Remark 5.2. In Theorem 5.2, we assume the differentiability of the coefficients g, b, σ and l, and the control takes values in a convex set $U \subset \mathbb{R}^m$. By using the spike variational arguments, we may allow the control to be valued in a non-convex set $U \subset \mathbb{R}^m$ and it is possible to use the relaxed controls avoiding differentiability assumption with respect to the control variable (see [1, Theorem 8.3.5 and Corollary 8.3.7, p. 276]). For the general cases with general cost functional and non-convex control set, the optimal control problem in this paper should be discussed further in the future.

6 Two related topics

6.1 Connections with Bismut's method for a particular example

Maximum principle for controlled stochastic differential equations (SDEs) was first discussed by Bismut [4–6], in which the adjoint equation is a linear backward stochastic differential equations (BSDEs). We compare the maximum principle herein with that in [6], and show the explicit relations between them.

When $g \equiv 0$ and $\delta = 0$, the controlled NSFDE (1.2) is reduced to the following controlled SDE,

$$\begin{cases} dX(t) = b(t, X(t), u(t))dt + \sigma(t, X(t), u(t))dW(t), & t \in [0, T], \\ X(0) = x, \end{cases}$$

and the cost functional is reduced to

$$J(u(\cdot)) = E\left[\int_0^T l(t, X(t), u(t))dt\right].$$

Suppose that (A1) and (A2) still hold. Assumption (A3) holds naturally in this case. The admissible control set and the optimal control problem are the same as in Section 2.

As a corollary of Theorem 5.2, we have the maximum principle.

Corollary 6.1. *Let $(\bar{X}(\cdot), \bar{u}(\cdot))$ be the optimal pair. Then there is $(Y, Z) \in \mathscr{M}^2(0, T)$ being the unique M-solution to the following equation*:

$$Y(t) = \bar{l}_x(t) + \int_t^T (\bar{b}'_x(t)Y(s) + \langle \bar{\sigma}_x(t),\, Z(s,t)\rangle)ds + \int_t^T Z(t,s)dW(s), \tag{6.1}$$

such that the following maximum condition holds:

$$\left\langle \bar{l}_u(t) + \bar{b}'_u(t)E_t\left[\int_t^T Y(s)ds\right] + \int_t^T \langle \bar{\sigma}_u(t),\, Z(s,t)\rangle ds,\, u - \bar{u}(t) \right\rangle \geqslant 0, \quad \forall\, u \in U, \quad dP(\omega) \otimes dt\text{-a.e.}.$$

Recall the maximum principle in Bismut [6].

Proposition 6.2. *Let $(\bar{X}(\cdot), \bar{u}(\cdot))$ be the optimal pair. Then there exists $(P, Q) \in \mathscr{S}^2_{\mathbb{F}}([0, T]; \mathbb{R}^n) \times \mathscr{L}^2_{\mathbb{F}}(0, T; \mathbb{R}^{n\times d})$ being the unique solution to the following BSDE*:

$$P(t) = \int_t^T [\bar{b}'_x(s)P(s) + \langle \bar{\sigma}_x(s),\, Q(s)\rangle + \bar{l}'_x(s)]ds - \int_t^T Q(s)dW(s), \tag{6.2}$$

such that

$$\langle \bar{l}_u(t) + \bar{b}'_u(t)P(t) + \langle \bar{\sigma}_u(t),\, Q(t)\rangle,\, u - \bar{u}(t)\rangle \geqslant 0, \quad \forall\, u \in U, \quad dP(\omega) \otimes dt\text{-a.e.}.$$

In fact, the two maximum principles possess the following relationship.

Theorem 6.3. *Let (Y,Z) and (P,Q) be as above, then*

$$P(t)=E_t\left[\int_t^T Y(s)ds\right],\quad \text{and}\quad Q(t)=\int_t^T Z(s,t)ds.$$

Moreover, the maximum principles in Corollary 6.1 *and Proposition* 6.2 *are equivalent.*

Proof. Similar to the proof in Theorem 5.2, we have $\frac{X_\varepsilon(\cdot)-\bar X(\cdot)}{\varepsilon}$ converges to $\chi(\cdot)$ in $\mathscr{S}^2_{\mathbb{F}}([0,T];\mathbb{R}^n)$, where $\chi(\cdot)$ satisfies

$$\chi(t)=\int_0^t[\bar b_x(s)\chi(s)+\bar b_u(s)\bar v(s)]ds+\int_0^t[\bar\sigma_x(s)\chi(s)+\bar\sigma_u(s)\bar v(s)]dW(s).$$

Denote

$$\rho(t):=\int_0^t\bar b_u(s)\bar v(s)ds+\int_0^t\bar\sigma_u(s)\bar v(s)dW(s).$$

The duality between linear SDE and BSDE (6.2) shows

$$E\int_0^T\langle\chi(t),\,\bar l_x(t)\rangle dt=E\int_0^T[\langle P(t),\,\bar b_u(t)\bar v(t)\rangle+\langle Q(t),\,\bar\sigma_u(t)\bar v(t)\rangle]dt,\tag{6.3}$$

and the duality between linear SDE and VNBSFE (6.1) shows

$$\begin{aligned}E\int_0^T\langle\chi(t),\,\bar l_x(t)\rangle dt&=E\int_0^T\langle\rho(t),\,Y(t)\rangle dt\\&=E\int_0^T\left\langle\int_0^t\bar b_u(s)\bar v(s)ds,\,Y(t)\right\rangle dt+E\int_0^T\left\langle\int_0^t\bar\sigma_u(s)\bar v(s)dW(s),\,Y(t)\right\rangle dt\\&=E\int_0^T\left\langle\bar b_u(t)\bar v(t),\,\int_t^T Y(s)ds\right\rangle dt+E\int_0^T\left\langle\bar\sigma_u(t)\bar v(t),\,\int_t^T Z(s,t)ds\right\rangle dt.\end{aligned}\tag{6.4}$$

Comparing (6.3) and (6.4), we prove the conclusion. □

Remark 6.1. From the foregoing discussion, the method in this paper dealing with the optimal control problem of NSFDEs is consistent with the traditional one dealing with SDEs. However, when the state equation of the optimal problem behaves more generally than semi-martingales, the traditional one is no longer applicable and for control problem (1.1) driven by NSFDEs (1.2), we have to appeal to the dual analysis of Section 4.

6.2 A sufficient condition and the application to an LQ problem

First, we have the following sufficient condition for optimal control with convex Hamiltonian functions.

Theorem 6.4. *Under Assumptions* (A1)–(A3)*, suppose that $H(t,\cdot,\cdot;P,Q)$ is convex and that $g(t,\cdot)$ is a linear functional. Let $(\bar X(\cdot),\bar u(\cdot))$ be an admissible pair. If for any $u\in U$,*

$$H(t,\bar X^t,u;Y,Z)\geqslant H(t,\bar X^t,\bar u(t);Y,Z),\quad dP(\omega)\otimes dt\text{-a.e.},$$

where (Y,Z) is the M-solution to VNBSFE (5.5) *associated with$(\bar X(\cdot),\bar u(\cdot))$, then $(\bar X(\cdot),\bar u(\cdot))$ is an optimal pair.*

Sketched proof. In view of the convexity of U, we have for any $u\in U$,

$$H(t,\bar X^t,\bar u(t)+\lambda(u-\bar u(t));Y,Z)\geqslant H(t,\bar X^t,\bar u(t);Y,Z).$$

Then

$$\lim_{\lambda\downarrow 0}\frac{H(t,\bar X^t,\bar u(t)+\lambda(u-\bar u(t));Y,Z)-H(t,\bar X^t,\bar u(t);Y,Z)}{\lambda}\geqslant 0,$$

i.e.,

$$\left\langle \bar{l}_u(t)+\bar{b}'_u(t)E_t\left[\int_t^T Y(s)ds\right]+\int_t^T\langle\bar{\sigma}_u(t),\ Z(s,t)\rangle ds,\ u-\bar{u}(t)\right\rangle \geqslant 0,\quad dP(\omega)\otimes dt\text{-a.e..} \tag{6.5}$$

For any admissible pair $(X(\cdot),u(\cdot))$, by convexity of H, we have

$$H(t,X^t,u(t);Y,Z)\geqslant H(t,\bar{X}^t,\bar{u}(t);Y,Z)+\bar{H}_x(t)(X^t-\bar{X}^t)+\bar{H}_u(t)(u(t)-\bar{u}(t)), \tag{6.6}$$

where $\bar{H}_x(t)=H_x(t,\bar{X}^t,\bar{u}(t);Y,Z)$, $\bar{H}_u(t)=H_u(t,\bar{X}^t,\bar{u}(t);Y,Z)$.

Note that, by (6.5) we have $\bar{H}_u(t)(u(t)-\bar{u}(t))\geqslant 0$. Integrating (6.6) on both sides, we obtain

$$E\int_0^T l(t,X^t,u(t))\geqslant E\int_0^T l(t,\bar{X}^t,\bar{u}(t))+\Gamma_1+\Gamma_2, \tag{6.7}$$

where

$$\begin{aligned}\Gamma_1&=E\left[\int_0^T(b'(t,\bar{X}^t,\bar{u}(t))-b'(t,X^t,u(t)))E_t\left[\int_t^T Y(s)ds\right]dt\right]\\&\quad+E\left[\int_0^T\left\langle\sigma(t,\bar{X}^t,\bar{u}(t))-\sigma(t,X^t,u(t)),\ \int_t^T Z(s,t)ds\right\rangle dt\right],\\\Gamma_2&=E\int_0^T\left\{\bar{l}_x(t)+\bar{b}'_x(t)E_t\left[\int_t^T Y(s)ds\right]+\left\langle\bar{\sigma}_x(t),\ \int_t^T Z(s,t)ds\right\rangle\right\}(X^t-\bar{X}^t)dt.\end{aligned}$$

From the relation

$$Y(t)=E[Y(t)]+\int_0^t Z(t,s)dW(s),$$

it follows that

$$\Gamma_1=-E\int_0^T\langle Y(t),\ X(t)-\bar{X}(t)-[g(t,X^t)-g(t,\bar{X}^t)]\rangle dt.$$

Directly substituting VNBSFE (5.5) into Γ_2, we get

$$\Gamma_2=E\int_0^T\left\langle Y(t),\ X(t)-\bar{X}(t)-\int_{-\delta}^0\bar{G}(t,r)[X(t+r)-\bar{X}(t+r)]dr\right\rangle dt.$$

Since $g(t,\cdot)$ is linear, in view of Lemma 2.1 and Assumption (A3), we have $\Gamma_1+\Gamma_2=0$, which together with ralation (6.7) implies the optimality of $(\bar{X}(\cdot),\bar{u}(\cdot))$. □

An immediate application of the sufficient condition of Theorem 6.4 and the maximum principle of Theorem 5.2 is the following LQ problem.

Example 6.5. *Consider the following LQ problem*:

$$\min_{u(\cdot)\in\mathcal{U}_{ad}} E\int_0^T(Q|u(t)|^2+S|X(t)|^2)dt$$

subject to

$$\begin{cases}d\left[X(t)-\dfrac{A}{\delta}\displaystyle\int_{-\delta}^0 X(t+r)\,dr-M_t\right]\\\quad=\left(\dfrac{B_1}{\delta}\displaystyle\int_{-\delta}^0 X(t+r)dr+D_1u(t)\right)dt+\left(\dfrac{B_2}{\delta}\displaystyle\int_{-\delta}^0 X(t+r)dr+D_2u(t)\right)dW(t),\quad t\in[0,T],\\X(t)=\phi(t),\quad t\in[-\delta,0].\end{cases} \tag{6.8}$$

For simplicity we take the dimension $d=m=n=1$ and the admissible control set $U=\mathbb{R}$. Here, we assume that $Q\in(0,\infty)$, $S\in[0,\infty]$, $A\in[0,1)$, $B_1,B_2,D_1,D_2\in\mathbb{R}$, and $(M_t)_{t\in[0,T]}$ is a bounded *continuous process unnecessarily being a semimartingale. From Theorems* 5.2 *and* 6.4, *it follows that there exists a unique optimal pair* $(\bar{X}(\cdot), \bar{u}(\cdot))$ *with*

$$\bar{u}(t) = -(2Q)^{-1}\left(D_1 E_t\left[\int_t^T Y(s)ds\right] + D_2\int_t^T Z(s,t)ds\right), \quad dP(\omega)\otimes dt\text{-}a.e.,$$

$\bar{X}$ *being the solution to NSFDE* (6.8) *associated with* $\bar{u}$ *and* (Y, Z) *satisfying VNBSFE*:

$$\begin{cases} Y(t) - AE_t\left[\dfrac{1}{\delta}\displaystyle\int_{-\delta}^0 Y(t-r)dr\right] = \displaystyle\int_t^T E_s\left[\dfrac{1}{\delta}\displaystyle\int_{-\delta}^0 B_1 Y(s-r)dr + \dfrac{1}{\delta}\displaystyle\int_{-\delta}^0 B_2 Z(s-r,t-r)dr\right]ds \\ \qquad\qquad\qquad\qquad\qquad - \displaystyle\int_t^T Z(t,s)dW(s) + 2S\bar{X}(t), \quad t\in[0,T], \\ Y(t) = 0, \quad t\in(T, T+\delta]. \end{cases} \tag{6.9}$$

Acknowledgements The author thanks the anonymous referees for their valuable comments and suggestions on the original manuscript of this work.

References

1 Ahmed N U. Dynamic Systems and Control with Applications. New Jersey-London-Beijing-Singapore: World Scientific, 2006

2 Ahmed N U. Deterministic and stochastic neutral systems on Banach spaces and their optimal feedback controls. J Nonlinear Syst Appl, 2013, 4: 1–10

3 Banks H, Kent G. Control of functional differential equations of retarded and neutral type to target sets in function space. SIAM J Control Optim, 1972, 10: 567–593

4 Bismut J. Théorie Probabiliste du Contrôle des Diffusions. Providence, RI: Amer Math Soc, 1976

5 Bismut J. Linear quadratic optimal stochastic control with random coefficients. SIAM J Control Optim, 1976, 14: 419–444

6 Bismut J. An introductory approach to duality in optimal stochastic control. SIAM Rev, 1978, 20: 62–78

7 Chen L, Wu Z. Maximum principle for stochastic optimal control problem of forward-backward system with delay. In: Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference. New York: IEEE, 2009, 2899–2904

8 Chen L, Wu Z. Maximum principle for the stochastic optimal control problem with delay and application. Automatica, 2010, 46: 1074–1080

9 Cushing C M. Integro-Differential Equations and Delay Models in Population Dynamics. Berlin: Springer-Verlag, 1977

10 El Karoui N, Peng S, Quenez M C. Backward stochastic differential equations in finance. Math Finance, 1997, 7: 1–71

11 Elsanosi I, Øksendal B, Sulem A. Some solvable stochastic control problems with delay. Stoch Stoch Rep, 2000, 71: 69–89

12 Gut A. An introduction to the theory of asymptotic martingales. In: Amarts and Set Function Processes. Lecture Notes in Mathematics, vol. 1042. Berlin-New York: Springer, 1983, 1–49

13 Hale J. Oscillations in neutral functional differential equations. In: Non-linear Mechanics. Rome: Edizioni Cremonese, 1973, 97–111

14 Hale J. Theory of Functional Differential Equations. New York: Springer-Verlag, 1977

15 Hale J K. Introduction to Functional Differential Equations. Berlin-New York: Springer-Verlag, 1993

16 Hu Y, Peng S. Maximum principle for optimal control of stochastic system of functional type. Stoch Anal Appl, 1996, 14: 283–301

17 Huang L, Mao X. Delay-dependent exponential stability of neutral stochastic delay systems. IEEE Trans Automat Control, 2009, 54: 147–152

18 Kent G. A maximum principle for optimal control problems with neutral functional differential systems. Bull Amer Math Soc, 1971, 77: 565–570

19 Kolmanovskii V, Khvilon E. Necessary conditions for optimal control of systems with deviating argument of neutral type. Autom Remote Control, 1969, 30: 327–339

20 Kolmanovskii V, Myshkis A. Introduction to the Theory and Applications of Functional Differential Equations. Boston, MA: Kluwer Academic Publisher, 1999

21 Kolmanovskii V B, Nosov V R. Stability and Periodic Modes of Control Systems with Aftereffect. Moscow: Nauka, 1981

22 Kolmanovskii V B, Nosov V R. Stability of Functional Differential Equations. London: Academic Press, 1986

23 Kolmanovskii V B, Shaikhet L. Construction of Lyapunov functionals for stochastic hereditary systems: A survey of some recent results. Math Comput Modelling, 2002, 36: 691–716

24 Lin J. Adapted solution of a backward stochastic nonlinear Volterra integral equation. Stoch Anal Appl, 2002, 20: 165–183

25 Liu K. Quadratic control problem of neutral Ornstein-Uhlenbeck processes with control delays. Discrete Contin Dyn Syst Ser B, 2013, 18: 1651–1661

26 Mao X. Exponential stability in mean square of neutral stochastic differential functional equations. Systems Control Lett, 1995, 26: 245–251

27 Øksendal B, Sulem A. A maximum principle for optimal control of stochastic systems with delay, with applications to finance. In: Optimal Control and Partial Differential Equation. Amsterdam: IOS Press, 2001, 64–79

28 Pardoux E, Peng S. Adapted solution of a backward stochastic differential equation. Systems Control Lett, 1990, 14: 55–61

29 Protter P. Stochastic Integration and Differential Equations: A New Approach. Berlin-New York-London-Paris-Tokyo-Hong Kong: Springer-Verlag, 1990

30 Ren Y, Xia N. Existence, uniqueness and stability of the solutions to neutral stochastic functional differential equations with infinite delay. Appl Math Comput, 2009, 210: 72–79

31 Salamon D. Control and Observation of Neutral Systems. London: Pitman Advanced Pub Program, 1984

32 Shaikhet L. Necessary and sufficient conditions of asymptotic mean square stability for stochastic linear difference equations. Appl Math Lett, 1997, 10: 111–115

33 Shaikhe L. Some new aspects of Lyapunov-type theorems for stochastic differential equations of neutral type. SIAM J Control Optim, 2010, 48: 4481–4499

34 Wei W. Neutral backward stochastic functional differential equations and their application. ArXiv:1301.3081, 2013

35 Yong J. Well-posedness and regularity of backward stochastic Volterra integral equations. Probab Theory Related Fields, 2008, 142: 21–77

36 Yu Z. The stochastic maximum principle for optimal control problems of delay systems involving continuous and impulse controls. Automatica, 2012, 48: 2420–2432