ON NEAR-OPTIMAL NECESSARY AND SUFFICIENT CONDITIONS FOR FORWARD-BACKWARD STOCHASTIC SYSTEMS WITH JUMPS, WITH APPLICATIONS TO FINANCE

Mokhtar Hafayed, Biskra, Petr Veverka, Praha, Syed Abbas, Mandi

(Received May 15, 2012)

Abstract. We establish necessary and sufficient conditions of near-optimality for nonlinear systems governed by forward-backward stochastic differential equations with controlled jump processes (FBSDEJs in short). The set of controls under consideration is necessarily convex. The proof of our result is based on Ekeland's variational principle and continuity in some sense of the state and adjoint processes with respect to the control variable. We prove that under an additional hypothesis, the near-maximum condition on the Hamiltonian function is a sufficient condition for near-optimality. At the end, as an application to finance, mean-variance portfolio selection mixed with a recursive utility optimization problem is given.

Keywords: stochastic near-optimal controls; jump processes; forward-backward stochastic systems with jumps; necessary and sufficient conditions for near-optimality; Ekeland's variational principle

MSC 2010: 60H10, 49K45, 93E20

1. Introduction

Let $T > 0$ be a fixed time horizon. We consider the stochastic control problem of near-optimality for systems governed by nonlinear FBSDEJs of the form

$$\begin{aligned}(1.1)\qquad \mathrm{d}x(t) &= f(t, x(t), u(t))\,\mathrm{d}t + \sigma(t, x(t), u(t))\,\mathrm{d}W(t)\\ &\quad + \int_{\Theta} c(t, x(t_-), u(t), \theta) N(\mathrm{d}\theta, \mathrm{d}t),\end{aligned}$$

The first author was partially supported by Algerian PNR project grant 08/u07/857, ATRST-ANDRU 2011-2013. The second author was supported by the Czech CTU grant SGS12/197/OHK4/3T/14 and MSMT grant INGO II INFRA LG12020.

$$
\begin{aligned}
-\mathrm{d}y(t) &= \int_\Theta g(t, x(t), y(t), z(t), r_t(\theta), u(t))\mu(\mathrm{d}\theta)\,\mathrm{d}t - z(t)\,\mathrm{d}W(t) \\
&\quad - \int_\Theta r_t(\theta) N(\mathrm{d}\theta, \mathrm{d}t), \\
x(0) &= \zeta, \quad y(T) = \varphi(x(T)).
\end{aligned}
$$

Here, $W = (W(t))_{t\in[0,T]}$ is a standard d-dimensional Brownian motion defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ satisfying the usual conditions. The initial condition ζ is an $\mathcal{F}_0$-measurable random variable. Further, $N(\mathrm{d}\theta, \mathrm{d}t)$ is a one dimensional Poisson martingale measure independent of W with characteristics $\mu(\mathrm{d}\theta)\,\mathrm{d}t$. The filtration $(\mathcal{F}_t)_{t\in[0,T]}$ is a canonical filtration of W and N augmented by $\mathbb{P}$-null sets.

The criterion to be minimized associated with the state equation (1.1) is defined by

$$
J(u(\cdot)) = \mathbb{E}\Big[\int_0^T \int_\Theta l(t, x(t), y(t), z(t), r_t(\theta), u(t))\mu(\mathrm{d}\theta)\,\mathrm{d}t + h(x(T)) + \gamma(y(0))\Big], \tag{1.2}
$$

and the value function is defined as

$$
V = \inf\{J(u(\cdot)), u(\cdot) \in \mathcal{U}_{\mathrm{ad}}\}, \tag{1.3}
$$

where f, σ, c, g, l, h, φ, γ are some appropriate functions and the process $u(\cdot)$ is a control from some set of admissible controls $\mathcal{U}_{\mathrm{ad}}$.

Near-optimization is as sensible and important as optimization both from the theory and application point of view. The stochastic control problems have been investigated extensively, both by Bellman's dynamic programming method [2] and by Pontryagin's maximum principle [20]. Many more near-optimal controls are available than the optimal ones. Indeed, optimal controls may not even exist in many situations, while near-optimal controls always exist. Various kinds of near-optimal control problems have been investigated in [6], [10], [12], [11], [13], [16], [18], [31], [32], [33]. In an interesting paper, Zhou [33] established second-order necessary as well as sufficient conditions for near-optimal stochastic controls for controlled diffusion, where the coefficients were assumed to be twice continuously differentiable. However, in Hafayed, Abbas and Veverka [11], the authors extended Zhou's maximum principle [33] to singular stochastic control. The near-optimal control problem for systems governed by Volterra integral equations has been studied in Pan and Teo [18]. The near-optimal stochastic control problem for systems governed by diffusions with jump processes, with application to finance has been investigated by Hafayed, Veverka and Abbas [12]. For justification of establishing a theory of near-optimal controls, see Zhou ([31], [32], [33] Introduction).

The stochastic maximum principle of optimality for FBSDEs has been studied by many authors, see e.g. [19], [25], [24], [23], [28], [29]. Necessary conditions of optimality for FBSDEs in global form, with uncontrolled diffusions coefficient was derived by Xu [28]. However, Shi and Wu [23] were the first who derived the stochastic maximum principle for the fully coupled forward-backward stochastic control system in global form. The near optimal control problems for FBSDEs have been treated in [1], [13]. Very recently, Yong [29] completely solved the problem of maximum principle of optimality for fully coupled FBSDEs. He considered an optimal control problem for general coupled FBSDEs with mixed initial-terminal conditions and derived necessary conditions for optimality when the control variable appears in the diffusion coefficients of the forward-equation and the control domain is not necessarily convex.

The stochastic optimal control problems for jump processes have been investigated by many authors, see e.g. [3], [5], [9], [17], [21]–[27]. Situ [26] first established the maximum principle for the stochastic control system with uncontrolled random jumps in global form. Tang and Li [27] completely proved the maximum principle in global form, where the control variable is allowed to enter both into the diffusion and jump coefficients by using the second-order expansion. Necessary and sufficient conditions of optimality for FBSDEJs were obtained by Shi and Wu [25], [24]. In an interesting paper, Shi [22] generalized Yong's maximum principle for FBSDEs obtained in Yong [29] to the jump case. He established the stochastic maximum principle for optimality for fully coupled FBSDEJs when the control variable appears both in the diffusion and jump coefficients and the control domain is not assumed to be convex. A good account and an extensive list of references on the stochastic maximum principle for FBSDEJs can be found in [17], [22].

Our purpose in this paper is to establish necessary as well as sufficient conditions for near-optimality for systems governed by nonlinear FBSDEJs. The control variable appears both in the diffusion and jump coefficients. The control domain is necessarily convex. The proof of our result is based on Ekeland's variational principle [6] and some delicate estimates of the state and adjoint processes. Moreover, we prove that under some additional assumptions, the necessary conditions are also sufficient for near-optimality. As an application to finance, the mean-variance portfolio selection mixed problem is provided.

This paper is organized as follows. In Section 2, we formulate the control problem and describe the assumptions of the model. In Sections 3 and 4, we establish the necessary and sufficient conditions of near-optimality. An application to finance is given in the last section.

2. Problem formulation and preliminaries

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a fixed filtered probability space equipped with a $\mathbb{P}$-completed right continuous filtration on which a d-dimensional Brownian motion $W = (W(t))_{t\in[0,T]}$ is defined. Let η be a one dimensional homogeneous $(\mathcal{F}_t)$-Poisson point process independent of W. We denote by $\widetilde{N}(\mathrm{d}\theta, \mathrm{d}t)$ the random counting measure induced by η, defined on $\Theta \times \mathbb{R}_+$, where Θ is a fixed nonempty subset of $\mathbb{R}^k$ with its Borel σ-field $\mathcal{B}(\Theta)$. Further, let $\mu(\mathrm{d}\theta)$ be the local characteristic measure of η, i.e., $\mu(\mathrm{d}\theta)$ is a σ-finite measure on $(\Theta, \mathcal{B}(\Theta))$ with $\mu(\Theta) < \infty$. We then define

$$N(\mathrm{d}\theta, \mathrm{d}t) = \widetilde{N}(\mathrm{d}\theta, \mathrm{d}t) - \mu(\mathrm{d}\theta)\,\mathrm{d}t,$$

where N is the Poisson martingale measure on $\mathcal{B}(\Theta) \times \mathcal{B}(\mathbb{R}_+)$ with local characteristics $\mu(\mathrm{d}\theta)\,\mathrm{d}t$. We assume that $(\mathcal{F}_t)_{t\in[0,T]}$ is the $\mathbb{P}$-augmentation of the natural filtration $(\mathcal{F}_t^{(W,N)})_{t\in[0,T]}$ defined as

$$\mathcal{F}_t^{(W,N)} = \sigma(W(s)\colon\ 0 \leqslant s \leqslant t) \vee \sigma\biggl(\int_0^s \int_B N(\mathrm{d}\theta, \mathrm{d}r)\colon\ 0 \leqslant s \leqslant t,\ B \in \mathcal{B}(\Theta)\biggr) \vee \mathcal{G},$$

where $\mathcal{G}$ denotes the totality of $\mathbb{P}$-null sets, and $\sigma_1 \vee \sigma_2$ denotes the σ-field generated by $\sigma_1 \cup \sigma_2$.

Notation. We will use the following notation in this paper:

1. Any element $a \in \mathbb{R}^n$ will be identified with a column vector whose i-th component is a_i. The norm here is defined as $|a| = |a_1| + \ldots + |a_n|$.

2. Denote by M^* the transpose of any vector or matrix M and denote the trace of the matrix M by $\mathrm{Tr}\{M\}$.

3. Denote by $\mathbb{E}$ the expectation with respect to $\mathbb{P}$.

4. For a function $\psi \in \mathcal{C}^1$ denote by ψ_x its gradient or Jacobian with respect to the variable x.

5. $\mathrm{sgn}(\cdot)$ denotes the sign function.

6. In the sequel, $\mathbb{L}^2_{\mathcal{F}}([0,T]; \mathbb{R}^n)$ denotes the Hilbert space of $(\mathcal{F}_t)$-progressively measurable, $\mathbb{R}^n$-valued processes $(x(t))_{t\in[0,T]}$ such that

$$\mathbb{E}\int_0^T |x(t)|^2\,\mathrm{d}t < \infty,$$

and $\mathbb{M}^2_{\mathcal{F}}([0,T]; \mathbb{R}^m)$ denotes the Hilbert space of $(\mathcal{F}_t)$-predictable, $\mathbb{R}^m$-valued processes $(\psi(t,\theta))_{t\in[0,T]}$ defined on $[0,T] \times \Theta$ such that

$$\mathbb{E}\int_0^T\int_\Theta |\psi(t,\theta)|^2 \mu(\theta)\,\mathrm{d}t < \infty.$$

7. Let $M \in \mathbb{R}^{n\times d\times n}$ and $N \in \mathbb{R}^{n\times d}$. Then the multiplication MN is a vector in $\mathbb{R}^n$ defined as $(MN)_i = \sum_{j=1}^{d}\sum_{l=1}^{n} M_{ijl}N_{il} = \sum_{j=1}^{d}\langle M_{ij\bullet}, N_{\bullet j}\rangle$, where $N_{\bullet j}$ denotes the j-th column of the matrix N.

8. C denotes a generic positive constant which may differ from line to line.

Definition 1. Let $T > 0$ be a fixed strictly positive real number and $\mathbb{U}$ a nonempty compact convex subset of $\mathbb{R}^m$. An admissible control is defined as a function $u(\cdot)\colon [0,T]\times\Omega \longrightarrow \mathbb{U}$ which is $(\mathcal{F}_t)$-predictable, $\mathbb{E}\int_0^T |u(t)|^2\,\mathrm{d}t < \infty$, and such that the equation (1.1) has a unique solution $u(\cdot) \in \mathcal{U}_{\mathrm{ad}}$.

Throughout this paper, we also assume that the coefficient functions

$$
\begin{aligned}
&f\colon\ [0,T]\times\mathbb{R}^n\times\mathbb{U}\to\mathbb{R}^n,\\
&\sigma\colon\ [0,T]\times\mathbb{R}^n\times\mathbb{U}\to\mathbb{R}^{n\times d},\\
&c\colon\ [0,T]\times\mathbb{R}^n\times\mathbb{U}\times\Theta\to\mathbb{R}^n,\\
&g\colon\ [0,T]\times\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^{m\times d}\times\mathbb{R}^m\times\mathbb{U}\to\mathbb{R}^m,\\
&l\colon\ [0,T]\times\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^{m\times d}\times\mathbb{R}^m\times\mathbb{U}\to\mathbb{R},\\
&h\colon\ \mathbb{R}^n\to\mathbb{R},\\
&\varphi\colon\ \mathbb{R}^n\to\mathbb{R}^m,\\
&\gamma\colon\ \mathbb{R}^m\to\mathbb{R},
\end{aligned}
$$

satisfy the following standing assumptions:

(H1) The functions f, σ, g, c, l, h, γ, φ are continuously differentiable in their variables including (x,y,z,r,u).

(H2) The derivatives f_x, f_u, σ_x, σ_u, φ_x, h_x, g_ϱ, l_ϱ $(\varrho = x,y,z,r,u)$ are bounded and

$$\int_\Theta (|c_x(t,x,u,\theta)|^2 + |c_u(t,x,u,\theta)|^2)\mu(\mathrm{d}\theta) < \infty.$$

Further, functions f, σ, c, c_x, c_u, h_x have at most linear growth in variable x, γ_y has at most linear growth in y, the function g has at most linear growth in x,y and $g_y = g_y(t,x,u)$, i.e., g_y is independent of y,z,r. For $\varrho = x,y,z,r,u$, the function l_ϱ has at most linear growth in x, y, z, r.

(H3) For $\varrho = x,u$ the functions f_ϱ, σ_ϱ, g_ϱ, g_y, c, c_ϱ, h, h_x, φ_x are globally Lipschitz in the variable x. The function γ is globally Lipschitz in the variable y. For $\varrho = x,u,y$ the function l_ϱ is globally Lipschitz in x, y, z, r.

From (H2) it follows that the function $g(t,x,\cdot,z,r,u)$ is "close to" linear function in y. Under the assumptions (H1), (H2) and (H3) equation (1.1) has a unique solution $(x(t), y(t), z(t), r_t(\cdot)) \in \mathbb{L}^2_{\mathcal{F}}([0,T];\mathbb{R}^n) \times \mathbb{L}^2_{\mathcal{F}}([0,T];\mathbb{R}^m) \times \mathbb{L}^2_{\mathcal{F}}([0,T];\mathbb{R}^{m\times d}) \times \mathbb{M}^2_{\mathcal{F}}([0,T];\mathbb{R}^m)$.

For any $u(\cdot) \in \mathcal{U}_{\mathrm{ad}}$ we introduce the following adjoint equations:

$$
(2.1)\qquad \begin{cases}
-\mathrm{d}p_t = \Big\{ f_x^*(t,x(t),u(t))p_t + \sigma_x^*(t,x(t),u(t))q_t \\
\qquad + \int_\Theta [c_x^*(t,x(t),u(t),\theta)R_t(\theta) + l_x^*(t,\Lambda_t(\theta),u(t))]\mu(\mathrm{d}\theta) \\
\qquad - \int_\Theta g_x^*(t,\Lambda_t(\theta),u(t))k_t\mu(\mathrm{d}\theta) \Big\}\,\mathrm{d}t - q_t\,\mathrm{d}W(t) \\
\qquad - \int_\Theta R_t(\theta)N(\mathrm{d}\theta,\mathrm{d}t), \\
p_T = -\varphi_x^*(x(T))k_T + h_x(x(T)), \\
\mathrm{d}k_t = \int_\Theta [g_y^*(t,\Lambda_t(\theta),u(t))k_t - l_y^*(t,\Lambda_t(\theta),u(t))]\mu(\mathrm{d}\theta)\,\mathrm{d}t \\
\qquad + \int_\Theta [g_z^*(t,\Lambda_t(\theta),u(t))k_t - l_z^*(t,\Lambda_t(\theta),u(t))]\mu(\mathrm{d}\theta)\,\mathrm{d}W(t) \\
\qquad + \int_\Theta [g_r^*(t,\Lambda_{t_-}(\theta),u(t))k_t - l_r^*(t,\Lambda_{t_-}(\theta),u(t))]N(\mathrm{d}\theta,\mathrm{d}t), \\
k_0 = -\gamma_y(y(0)),
\end{cases}
$$

where $\Lambda_t(\theta) = (x(t), y(t), z(t), r_t(\theta))$. Further, we define the Hamiltonian function $H : [0,T] \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{m\times d} \times \mathbb{R}^m \times \mathbb{U} \times \mathbb{R}^n \times \mathbb{R}^{n\times d} \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ associated with the stochastic control problem (1.1)–(1.2) as

$$
(2.2)\qquad H(t,x,y,z,r(\cdot),u,p,q,k,R(\cdot)) = p^* f(t,x,u) + \mathrm{Tr}\{q^*\sigma(t,x,u)\}
$$
$$
- \int_\Theta [k^* g(t,x,y,z,r(\theta),u) + R^*(\theta)c(t,x,u,\theta) - l(t,x,y,z,r(\theta),u)]\mu(\mathrm{d}\theta).
$$

Denoting $\Psi_t(\theta) = (p_t, q_t, k_t, R_t(\theta))$ and $\mathcal{H}(t,\cdot) \equiv H(t,\Lambda_t(\cdot),u_t,\Psi_t(\cdot))$, the adjoint equation (2.1) can be rewritten as follows:

$$
(2.3)\qquad \begin{cases}
-\mathrm{d}p_t = \mathcal{H}_x(t)\,\mathrm{d}t - q_t\,\mathrm{d}W(t) - \int_\Theta R_t(\theta)N(\mathrm{d}\theta,\mathrm{d}t), \\
p_T = -\varphi_x^*(x(T))k_T + h_x(x(T)), \\
\mathrm{d}k_t = -\mathcal{H}_y(t)\,\mathrm{d}t - \mathcal{H}_z(t)\,\mathrm{d}W(t) - \int_\Theta \mathcal{H}_r(t_-,\theta)N(\mathrm{d}\theta,\mathrm{d}t), \\
k_0 = -\gamma_y(y(0)).
\end{cases}
$$

It is a well known fact that under assumptions (H1)–(H3), the adjoint equations (2.1) or (2.3) admit a unique solution quartet $(p_t, q_t, k_t, R_t(\cdot))$ such that

$$
(p_t, q_t, k_t, R_t(\cdot)) \in \mathbb{L}^2_{\mathcal{F}}([0,T];\mathbb{R}^n) \times \mathbb{L}^2_{\mathcal{F}}([0,T];\mathbb{R}^{n\times d}) \times \mathbb{L}^2_{\mathcal{F}}([0,T];\mathbb{R}^m) \times \mathbb{M}^2_{\mathcal{F}}([0,T];\mathbb{R}^n).
$$

Moreover, since the corresponding derivatives of $f, \sigma, g, l, c, h, \varphi$ are bounded, we deduce by standard arguments that

$$\mathbb{E}\Big\{\sup_{t\in[0,T]} |p_t|^2 + \int_0^T |q_t|^2\,\mathrm{d}t + \int_0^T |k_t|^2\,\mathrm{d}t + \int_0^T\int_\Theta |R_t(\theta)|^2\mu(\mathrm{d}\theta)\,\mathrm{d}t\Big\} < \infty. \tag{2.4}$$

3. Necessary conditions for near-optimality of FBSDEJs

Our objective in this section is to derive near-optimality necessary conditions for FBSDEJs, where the control domain is necessarily convex. The proof of our main result is based on Ekeland's variational principle [6] and some estimates of the state and adjoint processes with respect to the control variable.

Let us recall the definition of the near-optimal control of order ε^λ as given in Zhou ([33], Definitions (2.1) and (2.2)) and Ekeland's variational principle which will be used in the sequel.

Definition 2 (Near-optimal control of order ε^λ)**.** For a given $\varepsilon > 0$ an admissible control $u^\varepsilon(\cdot)$ is called the near-optimal if

$$|J(u^\varepsilon(\cdot)) - V| \leqslant \mathcal{Q}(\varepsilon), \tag{3.1}$$

where $\mathcal{Q}(\cdot)$ is a function of ε satisfying $\lim_{\varepsilon\to 0}\mathcal{Q}(\varepsilon) = 0$. The estimator $\mathcal{Q}(\varepsilon)$ is called an error bound. If $\mathcal{Q}(\varepsilon) = C\varepsilon^\lambda$ for some $\lambda > 0$ independent of the constant C then $u^\varepsilon(\cdot)$ is called the near-optimal control of order ε^λ. If $\mathcal{Q}(\varepsilon) = \varepsilon$, the admissible control $u^\varepsilon(\cdot)$ is called ε-optimal.

Lemma 1 (Ekeland's Variational Principle [6])**.** *Let (E, d) be a complete metric space and $f\colon\ E \to \overline{\mathbb{R}}$ a lower semi-continuous function which is bounded from below. For a given $\varepsilon > 0$, suppose that there is $u^\varepsilon \in E$ satisfying*

$$f(u^\varepsilon) \leqslant \inf_{u\in E}(f(u)) + \varepsilon.$$

Then for any $\delta > 0$ there exists $u^\delta \in E$ such that

1. $f(u^\delta) \leqslant f(u^\varepsilon)$,
2. $d(u^\delta, u^\varepsilon) \leqslant \delta$,
3. $f(u^\delta) \leqslant f(u) + \varepsilon\delta^{-1}d(u, u^\delta)$ *for all* $u \in E$.

To apply Ekeland's variational principle to our problem, we must define a metric d on the space of admissible controls such that $(\mathcal{U}_{\mathrm{ad}}, d)$ becomes a complete metric space. For any $u(\cdot)$, $v(\cdot) \in \mathcal{U}_{\mathrm{ad}}$ we define

$$d(u(\cdot), v(\cdot)) = \mathbb{P}\otimes\mathrm{d}t\{(\omega, t) \in \Omega\times[0, T]\colon\ u(\omega, t) \neq v(\omega, t)\}, \tag{3.2}$$

where $\mathbb{P}\otimes \mathrm{d}t$ is the product measure of $\mathbb{P}$ with the Lebesgue measure $\mathrm{d}t$ on $[0,T]$. Moreover, it has been shown in the book by Yong and Zhou ([30], Lemma 6.4, 146–147) that

1. $(\mathcal{U}_{\rm ad}, d)$ is a complete metric space,
2. the cost function J is continuous from $\mathcal{U}_{\rm ad}$ into $\mathbb{R}$.

In the sequel, we adopt the following notation. For $u(\cdot)\in\mathcal{U}_{\rm ad}$ we denote by $\Lambda_t^u(\theta)=(x^u(t),y^u(t),z^u(t),\ r_t^u(\theta))$ and $\Psi_t^u(\theta)=(p_t^u,q_t^u,k_t^u,R_t^u(\theta))$ the solutions to state equation (1.1) and to adjoint equations (2.3) respectively, corresponding to $u(\cdot)$.

Lemma 2. *For any $\lambda\in[0,1/2)$ and for any $\varepsilon>0$ there exist $\bar u^\varepsilon(\cdot)\in\mathcal{U}_{\rm ad}$ and an $(\mathcal{F}_t)$-adapted process $(\bar p_t^\varepsilon,\bar q_t^\varepsilon,\bar k_t^\varepsilon,\overline{R}_t^\varepsilon(\cdot))$ such that for all $u\in\mathbb{U}$:*

$$\mathbb{E}\int_0^T H_u^*(t,\overline{\Lambda}_t^\varepsilon(\cdot),\bar u_t^\varepsilon,\overline{\Psi}_t^\varepsilon(\cdot))(u-\bar u^\varepsilon(t))\,\mathrm{d}t\geqslant -C\varepsilon^\lambda, \tag{3.3}$$

where $C=C(\lambda,\mu(\Theta),T)$ is a positive constant.

Proof. Applying Ekeland's variational principle with $\delta=\varepsilon^{1/2}$ there exists an admissible control $\bar u^\varepsilon(\cdot)$ such that

$$d(\bar u^\varepsilon(\cdot),u^\varepsilon(\cdot))\leqslant \varepsilon^{1/2}, \tag{3.4}$$

and $J^\varepsilon(\bar u^\varepsilon(\cdot))\leqslant J^\varepsilon(u(\cdot))$ for any $u(\cdot)\in\mathcal{U}_{\rm ad}$, where

$$J^\varepsilon(u(\cdot))=J(u(\cdot))+\varepsilon^{1/2}d(\bar u^\varepsilon(\cdot),u(\cdot)). \tag{3.5}$$

Notice that $\bar u^\varepsilon(\cdot)$ is near-optimal for the initial cost J and it is optimal for the new cost J^ε defined by (3.5).

Let $u^{\varepsilon,\varrho}(\cdot)$ denote a family of perturbed controls indexed by $\varrho\in[0,1]$ given by

$$u^{\varepsilon,\varrho}(t)=\bar u^\varepsilon(t)+\varrho(u(t)-\bar u^\varepsilon(t)).$$

By using the fact that $J^\varepsilon(\bar u^\varepsilon(\cdot))\leqslant J^\varepsilon(u^{\varepsilon,\varrho}(\cdot))$ and $d(\bar u^\varepsilon(\cdot),u^{\varepsilon,\varrho}(\cdot))\leqslant C\varrho$, we obtain

$$J(u^{\varepsilon,\varrho}(\cdot))-J(\bar u^\varepsilon(\cdot))\geqslant -\varepsilon^{1/2}d(\bar u^\varepsilon(\cdot),u^{\varepsilon,\varrho}(\cdot))\geqslant -C\varepsilon^{1/2}\varrho. \tag{3.6}$$

Dividing (3.6) by ϱ and sending ϱ to zero, we have

$$\frac{\mathrm{d}}{\mathrm{d}\varrho}(J(u^{\varepsilon,\varrho}(\cdot)))\Big|_{\varrho=0}\geqslant -C\varepsilon^{1/2}\geqslant -C\varepsilon^\lambda. \tag{3.7}$$

Finally, arguing as in Shi and Wu ([25] Theorem 2.1) for the left-hand side of inequality (3.7), the desired result follows. □

Now we are able to derive necessary conditions of near-optimality for systems governed by FBSDEJs, which is the main result of this paper.

Theorem 1 (Necessary Near-Optimal Maximum Principle). *Let the assumptions* (H1), (H2) *and* (H3) *hold. Then for any* $\lambda \in [0, 1/2)$ *there exists a positive constant* $C = C(\lambda, \mu(\Theta), T)$ *such that for any* $\varepsilon > 0$ *and any near-optimal control* $u^\varepsilon(\cdot)$, *the following inequality holds for all* $u \in \mathbb{U}$:

$$\mathbb{E}\int_0^T H_u^*(t, \Lambda_t^\varepsilon(\cdot), u^\varepsilon(t), \Psi_t^\varepsilon(\cdot))(u - u^\varepsilon(t))\,\mathrm{d}t \geqslant -C\varepsilon^\lambda. \tag{3.8}$$

To prove the above theorem, we need the following auxiliary results on the variation of the state and adjoint processes with respect to the control variable. First, let us recall the following proposition, which will be used to prove Lemma 3.

Proposition 1. *Let* $\mathcal{A}$ *be a predictable* σ*-field on* $\Omega \times [0, T]$, *and* f *an* $\mathcal{A} \otimes \mathcal{B}(\Theta)$*-measurable function such that*

$$\mathbb{E}\bigg\{\int_0^T\int_\Theta |f(s,\theta)|^2\mu(\mathrm{d}\theta)\,\mathrm{d}s\bigg\} < \infty.$$

Then for all $\gamma \geqslant 2$ *there exists a positive constant* $C = C(\gamma, T, \mu(\Theta))$ *such that*

$$\mathbb{E}\bigg\{\sup_{t\in[0,T]}\bigg|\int_0^t\int_\Theta f(s,\theta)N(\mathrm{d}s,\mathrm{d}\theta)\bigg|^\gamma\bigg\} \leqslant C\mathbb{E}\bigg\{\int_0^T\int_\Theta |f(s,\theta)|^\gamma\mu(\mathrm{d}\theta)\,\mathrm{d}s\bigg\}.$$

Proof. See Bouchard and Elie ([4], Appendix). □

The next lemma is an extension of Lemma 3 in Zhou [33], to the forward-backward stochastic differential equations.

Lemma 3. *Let* $(x^u(t), y^u(t), z^u(t), r^u(t,\cdot))_{t\in[0,T]}$ *and* $(x^v(t), y^v(t), z^v(t), r^v(t,\cdot))_{t\in[0,T]}$ *be two solutions of the state equation* (1.1) *associated, respectively, with some admissible controls* $u(\cdot)$ *and* $v(\cdot)$ *such that* $d(u(\cdot), v(\cdot)) \leqslant 1$. *Then for any* $\alpha \in (0,1)$, $\beta \geqslant 0$ *and* $\beta' \in (1,2]$ *satisfying* $\alpha\beta < 1$, $\alpha\beta' < 1$ *there exist positive constants* $C_1 = C_1(\alpha, \beta, \mu(\Theta), T)$ *and* $C_2 = C_2(\alpha, \beta, \beta', \mu(\Theta), T)$ *such that*

$$\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}|x^u(t) - x^v(t)|^\beta\Big) \leqslant C_1 d(u(\cdot), v(\cdot))^{\alpha\beta/2}, \tag{3.9}$$

$$\begin{aligned}\mathbb{E}\bigg\{\sup_{0\leqslant t\leqslant T}|y^u(t) - y^v(t)|^{\beta'} + \int_0^T |z^u(t) - z^v(t)|^{\beta'}\,\mathrm{d}t \\ + \int_0^T\int_\Theta |r_s^u(\theta) - r_s^v(\theta)|^{\beta'}\mu(\mathrm{d}\theta)\,\mathrm{d}s\bigg\} \leqslant C_2 d(u(\cdot), v(\cdot))^{\alpha\beta'/2}.\end{aligned} \tag{3.10}$$

P r o o f. First we show that under (H1)–(H3) the solution process $(y(t))_{t\in[0,T]}$ to Equation (1.1) corresponding to the backward component satisfies

$$\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}|y(t)|^p\Big)<\infty\ \forall\, p\geqslant 0. \tag{3.11}$$

To prove it, denote $\Lambda_t(\theta)=(x(t),y(t),z(t),r_t(\theta))$ and first assume $p\geqslant 2$. Then we have for each $t\in[0,T]$

$$y(t)=\mathbb{E}\Big[\varphi(x(T))+\int_t^T\int_\Theta g(s,\Lambda_s(\theta),u(s))\mu(\mathrm{d}\theta)\,\mathrm{d}s\Big|\mathcal{F}_t\Big].$$

Now by the Jensen and generalized triangle inequality we derive

$$|y(t)|^p\leqslant C\mathbb{E}\Big[|\varphi(x(T))|^p+\int_t^T\int_\Theta |g(s,\Lambda_s(\theta),u(s))|^p\mu(\mathrm{d}\theta)\,\mathrm{d}s\Big|\mathcal{F}_t\Big].$$

Due to the Lipschitz property of φ and at most linear growth of g in x and y we obtain

$$|y(t)|^p\leqslant C\mathbb{E}\Big[1+|x(T)|^p+\int_t^T(|x(s)|^p+|y(s)|^p)\,\mathrm{d}s\Big|\mathcal{F}_t\Big], \tag{3.12}$$

and hence,

$$\mathbb{E}|y(t)|^p\leqslant C\mathbb{E}\Big[1+|x(T)|^p+\int_t^T(|x(s)|^p+|y(s)|^p)\,\mathrm{d}s\Big]. \tag{3.13}$$

By the Gronwall inequality and the fact that $\mathbb{E}\big(\sup_{0\leqslant t\leqslant T}|x(t)|^p\big)<\infty$, $p\geqslant 0$, we can show that

$$\sup_{0\leqslant t\leqslant T}\mathbb{E}|y(t)|^p<\infty. \tag{3.14}$$

To obtain a similar estimate with the supremum inside the expectation we realize that due to (3.12)

$$\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}|y(t)|^p\Big)\leqslant C+C\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}\mathbb{E}\Big[|x(T)|^p+\int_t^T(|x(s)|^p+|y(s)|^p)\,\mathrm{d}s\Big|\mathcal{F}_t\Big]\Big).$$

By virtue of the fact that $\mathbb{E}\big(\sup_{0\leqslant t\leqslant T}|x(t)|^p\big)<\infty$, $p\geqslant 0$, and (3.14) the processes

$$M_1(t)=\mathbb{E}[|x(T)|^p|\mathcal{F}_t],\qquad M_2(t)=\mathbb{E}\Big[\int_0^T|x(s)|^p\,\mathrm{d}s\Big|\mathcal{F}_t\Big],$$

and

$$M_3(t) = \mathbb{E}\Big[\int_0^T |y(s)|^p \, \mathrm{d}s \Big| \mathcal{F}_t \Big], \quad t \in [0,T],$$

are $\mathbb{L}^2$-martingales with càdlàg paths and therefore, the Burkholder-Davis-Gundy and Cauchy-Schwarz inequalities imply

$$\mathbb{E}\Big(\sup_{0 \leqslant t \leqslant T} |M_1(t)| \Big) \leqslant C \mathbb{E} \sqrt{\langle M_1 \rangle_T} \leqslant C \sqrt{\mathbb{E} \langle M_1 \rangle_T} < \infty. \tag{3.15}$$

As for M_2 (and similarly for M_3) we take into account that

$$\begin{aligned} \mathbb{E}\Big(\sup_{0 \leqslant t \leqslant T} \mathbb{E}\Big[\int_t^T |x(s)|^p \, \mathrm{d}s \Big| \mathcal{F}_t \Big] \Big) & \\ \leqslant \mathbb{E}\Big(\sup_{0 \leqslant t \leqslant T} |M_2(t)| \Big) + \mathbb{E}|M_2(0)| & \\ \leqslant 2\mathbb{E}\Big(\sup_{0 \leqslant t \leqslant T} |M_2(t)| \Big), & \end{aligned} \tag{3.16}$$

and repeat the line from (3.15). Thus we conclude that indeed

$$\mathbb{E}\Big(\sup_{0 \leqslant t \leqslant T} |y(t)|^p \Big) < \infty, \quad p \geqslant 2.$$

The case when $p \in [0,2)$ follows easily by the Hölder inequality.

Next denote $\tilde{x}(t) = x^u(t) - x^v(t)$ and define similarly the processes $\tilde{y}(t)$, $\tilde{z}(t)$ and $\tilde{r}_t(\cdot)$. First we prove (3.9) for two cases. First, let $\beta = 2p$, $p \geqslant 1$. Then by using standard techniques we have that

$$\begin{aligned} \mathbb{E}\Big(\sup_{0 \leqslant t \leqslant T} |\tilde{x}(t)|^{2p} \Big) &\leqslant C \mathbb{E} \int_0^T |f(t, x^u(t), u(t)) - f(t, x^v(t), v(t))|^{2p} \, \mathrm{d}t \\ &+ C \mathbb{E} \int_0^T |\sigma(t, x^u(t), u(t)) - \sigma(t, x^v(t), v(t))|^{2p} \, \mathrm{d}t \\ &+ C \mathbb{E} \int_0^T \int_\Theta |c(t, x^u(t), u(t), \theta) - c(t, x^v(t), v(t), \theta)|^{2p} \mu(\mathrm{d}\theta) \, \mathrm{d}t \\ &= \mathbb{I}_1 + \mathbb{I}_2 + \mathbb{I}_3. \end{aligned}$$

By the definition of the metrics d and by the growth conditions on f we can estimate $\mathbb{I}_1$ in the following way:

$$\begin{aligned} \mathbb{I}_1 &\leqslant C \mathbb{E} \int_0^T |f(t, x^u(t), u(t)) - f(t, x^u(t), v(t))|^{2p} \, \mathrm{d}t \\ &\quad + C \mathbb{E} \int_0^T |f(t, x(t), v(t)) - f(t, x^v(t), v(t))|^{2p} \, \mathrm{d}t \\ &= \mathbb{I}_{1,1} + \mathbb{I}_{1,2}. \end{aligned}$$

Applying the Hölder inequality with $a = 1/(\alpha p) > 1$, $b > 1$ such that $1/a + 1/b = 1$ leads to

$$\begin{aligned}\mathbb{I}_{1,1} &= C\mathbb{E}\int_0^T |f(t, x^u(t), u(t)) - f(t, x^u(t), v(t))|^{2p}\mathbf{1}_{\{u(t)\neq v(t)\}}(t)\,\mathrm{d}t \\ &\leqslant C\Big(\mathbb{E}\int_0^T (1 + |x^u(t)|^{2pb})\,\mathrm{d}t\Big)^{1/b}(d(u(\cdot), v(\cdot)))^{\alpha p} \\ &\leqslant Cd(u(\cdot), v(\cdot))^{\alpha p}.\end{aligned}$$

The estimate of $\mathbb{I}_{1,2}$ follows from the Lipschitz property of f and from the fact that

$$\mathbb{I}_{1,2} \leqslant C\mathbb{E}\int_0^T |\tilde{x}(t)|^{2p}\,\mathrm{d}t \leqslant C\int_0^T \mathbb{E}\Big(\sup_{0\leqslant t\leqslant \tau}|\tilde{x}(t)|^{2p}\Big)\,\mathrm{d}\tau.$$

Repeating the same steps we derive similar estimates for $\mathbb{I}_2$ and $\mathbb{I}_3$ obtaining

$$\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}|\tilde{x}(t)|^{2p}\Big) \leqslant Cd(u(\cdot), v(\cdot))^{\alpha\beta/2} + C\int_0^T \mathbb{E}\Big(\sup_{0\leqslant t\leqslant \tau}|\tilde{x}(t)|^{2p}\Big)\,\mathrm{d}\tau.$$

Application of Gronwall's lemma therefore leads to

$$\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}|\tilde{x}(t)|^{2p}\Big) \leqslant Cd(u(\cdot), v(\cdot))^{\alpha\beta/2}. \tag{3.17}$$

Now, let $\beta = 2p$, $p \in [0, 1)$. Due to the Hölder inequality (with $a = 1/p > 1$) and the preceding result one has

$$\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}|\tilde{x}(t)|^{2p}\Big) \leqslant \Big(\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}|\tilde{x}(t)|^{2}\Big)\Big)^p \leqslant Cd(u(\cdot), v(\cdot))^{\alpha p}.$$

Now we proceed to proving the inequality (3.10). Denote $\Lambda_t^u(\theta) = (x^u(t), y^u(t), z^u(t), r_t^u(\theta))$ and consider again the two cases for β'.

First, let $\beta' = 2$. It is easy to see that the triple $(\tilde{y}(t), \tilde{z}(t), \tilde{r}_t(\theta))_{t\in[0,T]}$ satisfies

$$\begin{aligned}-\tilde{y}(t) - \int_t^T \tilde{z}(s)\,\mathrm{d}W(s) - \int_t^T\int_\Theta \tilde{r}_s(\theta)N(\mathrm{d}\theta, \mathrm{d}s) \\ = -(\varphi(x^u(T)) - \varphi(x^v(T))) \\ + \int_t^T\int_\Theta (g(s, \Lambda_s^u(\theta), u(s)) - g(s, \Lambda_s^v(\theta), v(s)))\mu(\mathrm{d}\theta)\,\mathrm{d}s.\end{aligned}$$

Now by scalar multiplying each side of the above equality with itself and taking expectation one obtains

$$
\begin{aligned}
(3.18)\qquad \mathbb{E}\bigg(|\tilde{y}(t)|^2 + \int_t^T |\tilde{z}(s)|^2\,\mathrm{d}s + \int_t^T\int_\Theta |\tilde{r}_s(\theta)|^2\mu(\mathrm{d}\theta)\,\mathrm{d}s\bigg)
&\leqslant C\mathbb{E}\bigg(\int_t^T\int_\Theta |g(s,\Lambda_s^u(\theta),u(s)) - g(s,\Lambda_s^v(\theta),v(s))|\mu(\mathrm{d}\theta)\,\mathrm{d}s\bigg)^2 \\
&\quad + C\mathbb{E}|\varphi(x^u(T)) - \varphi(x^v(T))|^2 \\
&= \mathbb{I}_1 + \mathbb{I}_2.
\end{aligned}
$$

We have used the facts that

$$
\mathbb{E}\bigg[\tilde{y}(t)\int_t^T \tilde{z}(s)\,\mathrm{d}W(s)\bigg] = 0, \qquad \mathbb{E}\bigg[\tilde{y}(t)\int_t^T\int_\Theta \tilde{r}_s(\theta)N(\mathrm{d}\theta,\mathrm{d}s)\bigg] = 0,
$$

$$
\mathbb{E}\bigg[\int_t^T\int_\Theta \tilde{r}_s(\theta)N(\mathrm{d}\theta,\mathrm{d}s)\int_t^T \tilde{z}(s)\,\mathrm{d}W(s)\bigg] = 0,
$$

due to independence of W and N and due to independence of its increments after time t on $\mathcal{F}_t$.

By the Lipschitz property of φ and by (3.9) the estimate of $\mathbb{I}_2$ follows easily as

$$
(3.19)\qquad \mathbb{I}_2 \leqslant C\mathbb{E}|\tilde{x}(T)|^2 \leqslant Cd(u(\cdot),v(\cdot))^\alpha.
$$

To estimate $\mathbb{I}_1$ let us write

$$
\begin{aligned}
\mathbb{I}_1 &\leqslant C\mathbb{E}\bigg(\int_t^T\int_\Theta |g(s,\Lambda_s^u(\theta),u(s)) - g(s,\Lambda_s^v(\theta),u(s))|^2\mu(\mathrm{d}\theta)\,\mathrm{d}s\bigg)^2 \\
&\quad + C\mathbb{E}\int_t^T\int_\Theta |g(s,\Lambda_s^v(\theta),u(s)) - g(s,\Lambda_s^v(\theta),v(s))|^2\mathbf{1}_{\{u(s)\neq v(s)\}}(s)\mu(\mathrm{d}\theta)\,\mathrm{d}s \\
&= \mathbb{I}_{1,1} + \mathbb{I}_{1,2}.
\end{aligned}
$$

By virtue of the growth condition on g in (H2) (recall that $|g(t,x,y,z,r)| \leqslant C(1 + |x| + |y|)$), the Hölder inequality with $a = 1/\alpha > 1$, $b > 1$ such that $1/a + 1/b = 1$ and by (3.11) it follows that

$$
\begin{aligned}
(3.20)\qquad \mathbb{I}_{1,2} &\leqslant C\mathbb{E}\int_t^T\int_\Theta (1 + \sup_{0\leqslant s\leqslant T}|x^v(s)|^2 + \sup_{0\leqslant s\leqslant T}|y^v(s)|^2)\mathbf{1}_{\{u(s)\neq v(s)\}}(s)\mu(\mathrm{d}\theta)\,\mathrm{d}s \\
&\leqslant C\Big(1 + \mathbb{E}\Big[\sup_{0\leqslant s\leqslant T}|x^v(s)|^{2/(1-\alpha)}\Big] + \mathbb{E}\Big[\sup_{0\leqslant s\leqslant T}|y^v(s)|^{2/(1-\alpha)}\Big]\Big)^{1-\alpha} \\
&\quad \times\bigg(\mathbb{E}\int_t^T\int_\Theta \mathbf{1}_{\{u(s)\neq v(s)\}}(s)\mu(\mathrm{d}\theta)\,\mathrm{d}s\bigg)^\alpha \\
&\leqslant Cd(u(\cdot),v(\cdot))^\alpha.
\end{aligned}
$$

Now, let us turn to estimate $\mathbb{I}_{1,1}$. Using the Lipschitz property of g and the Cauchy-Schwarz inequality one has

$$\begin{aligned}\mathbb{I}_{1,1} \leqslant CT\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}|\tilde{x}(t)|^2\Big) \\ + CT\mathbb{E}\int_t^T|\tilde{y}(s)|^2\,\mathrm{d}s + C(T-t)\mathbb{E}\int_t^T|\tilde{z}(s)|^2\,\mathrm{d}s \\ + C(T-t)\mathbb{E}\int_t^T\int_\Theta|\tilde{r}_s(\theta)|^2\mu(\mathrm{d}\theta)\,\mathrm{d}s.\end{aligned} \tag{3.21}$$

Let us now take a particular $t\in[0,T]$ so that $\delta=T-t<1/(2C)$. Then putting together (3.18), (3.19), (3.20), (3.21), and (3.9) we obtain after rearranging the terms the estimate

$$\begin{aligned}\mathbb{E}|\tilde{y}(t)|^2 + \frac{1}{2}\mathbb{E}\int_t^T|\tilde{z}(s)|^2\,\mathrm{d}s + \frac{1}{2}\mathbb{E}\int_t^T\int_\Theta|\tilde{r}_s(\theta)|^2\mu(\mathrm{d}\theta)\,\mathrm{d}s \\ \leqslant Cd(u(\cdot),v(\cdot))^\alpha\ \forall\, t\in[T-\delta,T].\end{aligned}$$

Repeating similar steps we obtain a similar estimate for all $t\in[T-2\delta,T-\delta]$ and after a finite number of iterations we cover the whole interval $[0,T]$.

The proof of the same estimate but for $\beta'\in[0,2)$ follows by the Hölder inequality similarly to the proof of (3.9). This completes the proof of Lemma 3. □

Lemma 4. *Let $(p_t^u,q_t^u,k_t^u,R_t^u(\cdot))$ and $(p_t^v,q_t^v,k_t^v,R_t^v(\cdot))$ be two adjoint solution processes to equations* (2.1) *corresponding to some admissible controls $u(\cdot)$ and $v(\cdot)$ respectively with $d(u(\cdot),v(\cdot))\leqslant 1$. Then for any $\beta\in(1,2)$ and $\alpha\in(0,1)$ satisfying $(1+\alpha)\beta<2$ there is a positive constant $C=C(\alpha,\beta,\mu(\Theta),T)$ such that*

$$\begin{aligned}\mathbb{E}\int_0^T\Big\{|p_t^u-p_t^v|^\beta+|q_t^u-q_t^v|^\beta+\int_\Theta|R_t^u(\theta)-R_t^v(\theta)|^\beta\mu(\mathrm{d}\theta)\Big\}\,\mathrm{d}t \\ \leqslant Cd(u(\cdot),v(\cdot))^{\alpha\beta/2}\end{aligned} \tag{3.22}$$

and

$$\mathbb{E}\int_0^T|k_t^u-k_t^v|^\beta\,\mathrm{d}t\leqslant Cd(u(\cdot),v(\cdot))^{\alpha\beta/2}. \tag{3.23}$$

Proof. For each $t\in[0,T]$ and $\theta\in\Theta$, we denote $\tilde{p}_t=p_t^u-p_t^v$, $\tilde{q}_t=q_t^u-q_t^v$, $\tilde{k}_t=k_t^u-k_t^v$ and $\widetilde{R}_t(\theta)=R_t^u(\theta)-R_t^v(\theta)$. Further, similarly to the previous proof denote $\Lambda_t^u(\theta)=(x^u(t),y^u(t),z^u(t),r_t^u(\theta))$.

We start by proving (3.23). Note that the process $(\tilde{k}_t)_{t\in[0,T]}$ satisfies the following (forward) stochastic differential equation:

$$\begin{cases} \mathrm{d}\tilde{k}_t = & \displaystyle\int_\Theta (g_y^*(t,\Lambda_t^u(\theta),u(t))\tilde{k}_t + G_y(t,\theta))\mu(\mathrm{d}\theta)\,\mathrm{d}t \\ & \displaystyle + \int_\Theta (g_z^*(t,\Lambda_t^u(\theta),u(t))\tilde{k}_t + G_z(t,\theta))\mu(\mathrm{d}\theta)\,\mathrm{d}W(t) \\ & \displaystyle + \int_\Theta (g_r^*(t,\Lambda_{t-}^u(\theta),u(t))\tilde{k}_t + G_r(t,\theta))N(\mathrm{d}\theta,\mathrm{d}t), \quad t\in(0,T], \\ \tilde{k}_0 = & -(\gamma_y(y^u(0)) - \gamma_y(y^v(0))), \end{cases}$$

where

$$\begin{aligned} G_y(t,\theta) &= [g_y(t,\Lambda_t^u(\theta),u(t)) - g_y(t,\Lambda_t^v(\theta),v(t))]k_t^v \\ &\quad + l_y(t,\Lambda_t^u(\theta),u(t)) - l_y(t,\Lambda_t^v(\theta),v(t)), \\ G_z(t,\theta) &= [g_z(t,\Lambda_t^u(\theta),u(t)) - g_z(t,\Lambda_t^v(\theta),v(t))]k_t^v \\ &\quad + l_z(t,\Lambda_t^u(\theta),u(t)) - l_z(t,\Lambda_t^v(\theta),v(t)), \\ G_r(t,\theta) &= [g_r(t,\Lambda_{t-}^u(\theta),u(t)) - g_r(t,\Lambda_{t-}^v(\theta),v(t))]k_t^v \\ &\quad + l_r(t,\Lambda_{t-}^u(\theta),u(t)) - l_r(t,\Lambda_{t-}^v(\theta),v(t)). \end{aligned}$$

Then we have for each $t\in[0,T]$

$$\begin{aligned} -\tilde{k}_t &= (\gamma_y(y^u(0)) - \gamma_y(y^v(0))) \\ &\quad + \int_0^t\int_\Theta (g_y^*(t,\Lambda_t^u(\theta),u(t))\tilde{k}_s + G_y(s,\theta))\mu(\mathrm{d}\theta)\,\mathrm{d}s \\ &\quad + \int_0^t\int_\Theta (g_z^*(t,\Lambda_t^u(\theta),u(t))\tilde{k}_s + G_z(s,\theta))\mu(\mathrm{d}\theta)\,\mathrm{d}W(s) \\ &\quad + \int_0^t\int_\Theta (g_r^*(t,\Lambda_{t-}^u(\theta),u(t))\tilde{k}_s + G_r(s,\theta))N(\mathrm{d}\theta,\mathrm{d}s). \end{aligned}$$

The adjoint equation for $\tilde{k}_t$ is the ODE

$$\begin{cases} \mathrm{d}\psi_t = -\left[\displaystyle\int_\Theta g_y^*(t,\Lambda_t^u(\theta),u(t))\psi_t\mu(\mathrm{d}\theta) + |\tilde{k}_t|^{\beta-1}\operatorname{sgn}(\tilde{k}_t)\right]\mathrm{d}t, \\ \psi_T = 0, \end{cases} \tag{3.24}$$

where $\beta\in(1,2)$. The equation has random coefficients (due to processes $(x(t))$ and $(u(t))$, see (H3)) but has neither the diffusion nor the jump part. Since g_y is bounded and Lipschitz in x, it can be easily shown that (3.24) admits a unique solution in

$\mathbb{L}^2$ which can be found pathwise. Moreover, the following estimate holds for $\gamma > 2$, $1/\beta + 1/\gamma = 1$:

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |\psi_t|^\gamma\Big] < \infty. \tag{3.25}$$

Indeed, by integrating (3.24) from t to T and taking $|\cdot|^\gamma$ one gets

$$\begin{aligned}|\psi_t|^\gamma &\leqslant C\bigg|\int_t^T \psi_s \int_\Theta g_y^*(t, \Lambda_t^u(\theta), u(t))\mu(\mathrm{d}\theta)\,\mathrm{d}s\bigg|^\gamma + C\bigg|\int_t^T |\tilde{k}_s|^{\beta-1}\operatorname{sgn}(\tilde{k}_s)\,\mathrm{d}s\bigg|^\gamma \\ &\leqslant C\bigg(\int_t^T |\psi_s| \int_\Theta |g_y^*(t, \Lambda_t^u(\theta), u(t))|\mu(\mathrm{d}\theta)\,\mathrm{d}s\bigg)^\gamma + Cm^{\gamma/2}\bigg(\int_t^T |\tilde{k}_s|^{\beta-1}\,\mathrm{d}s\bigg)^\gamma,\end{aligned}$$

where m is the dimension of $\tilde{k}_t$. Using now the Hölder inequality and boundedness of g_y by K, we have

$$\begin{aligned}|\psi_t|^\gamma &\leqslant CK^\gamma\mu^\gamma(\Theta)(T-t)^{\gamma-1}\int_t^T |\psi_s|^\gamma\,\mathrm{d}s + Cm^{\gamma/2}(T-t)^{\gamma-1}\int_t^T |\tilde{k}_s|^\beta\,\mathrm{d}s \\ &\leqslant \widetilde{C}\int_t^T |\psi_s|^\gamma\,\mathrm{d}s + \widetilde{C}\int_t^T |\tilde{k}_s|^\beta\,\mathrm{d}s,\end{aligned} \tag{3.26}$$

where $\widetilde{C} = \widetilde{C}(T, \gamma, K, \mu(\Theta), m)$.

Now, first taking expectation and supremum over time leads to

$$\sup_{t\in[0,T]} \mathbb{E}|\psi_t|^\gamma \leqslant \widetilde{C}\mathbb{E}\int_0^T |\psi_s|^\gamma\,\mathrm{d}s + \widetilde{C}\mathbb{E}\int_0^T |\tilde{k}_s|^\beta\,\mathrm{d}s. \tag{3.27}$$

Note that by the Hölder inequality one has

$$\mathbb{E}\int_0^T |\tilde{k}_s|^\beta\,\mathrm{d}s \leqslant C\bigg(\mathbb{E}\int_0^T |\tilde{k}_s|^2\,\mathrm{d}s\bigg)^{\beta/2} \leqslant \widetilde{K} < \infty. \tag{3.28}$$

Therefore, applying Gronwall's lemma to (3.27) and (3.28), we conclude that

$$\sup_{t\in[0,T]} \mathbb{E}|\psi_t|^\gamma \leqslant \widetilde{K}\exp(\widetilde{C}T) < \infty. \tag{3.29}$$

Now, taking the supremum over time in (3.26), taking expectation and applying the previous estimates, we end with

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in[0,T]} |\psi_t|^\gamma\Big] &\leqslant \widetilde{C}\mathbb{E}\int_0^T |\psi_s|^\gamma\,\mathrm{d}s + \widetilde{C}\mathbb{E}\int_0^T |\tilde{k}_s|^\beta\,\mathrm{d}s \\ &\leqslant \widetilde{C}\widetilde{K}\exp(\widetilde{C}T)T + \widetilde{C}\widetilde{K} < \infty.\end{aligned}$$

Let us derive now a better estimate of $\mathbb{E}\Big[\sup_{t\in[0,T]} |\psi_t|^\gamma\Big]$. Taking expectation in (3.26), applying Gronwall's inequality and taking supremum, one gets

$$\sup_{t\in[0,T]} \mathbb{E}|\psi_t|^\gamma \leqslant K\mathbb{E}\int_0^T |\tilde{k}_s|^\beta \,\mathrm{d}s. \tag{3.30}$$

Using the previous estimate, one can derive along similar lines the estimate

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |\psi_t|^\gamma\Big] \leqslant K\mathbb{E}\int_0^T |\tilde{k}_s|^\beta \,\mathrm{d}s. \tag{3.31}$$

The duality of ψ_t and $\tilde{k}_t$ is shown by the Itô formula applied to $\tilde{k}_t^*\psi_t$ on $[0,T]$. After taking $\mathbb{E}(\cdot)$ it gives

$$\mathbb{E}\int_0^T |\tilde{k}_t|^\beta \,\mathrm{d}t = -\mathbb{E}[\tilde{k}_0\psi_0] - \mathbb{E}\int_0^T\int_\Theta \psi_t^* G_y(t,\theta)\mu(\mathrm{d}\theta)\,\mathrm{d}t. \tag{3.32}$$

The estimation of $\mathbb{E}[\tilde{k}_0\psi_0]$ goes as follows (using the Hölder inequality for $\gamma > 2$ and the uniform estimate (3.31) for ψ)

$$\begin{aligned}
\mathbb{E}[\tilde{k}_0\psi_0] &\leqslant (\mathbb{E}|\psi_0|^\gamma)^{1/\gamma}(\mathbb{E}|\tilde{k}_0|^\beta)^{1/\beta}\\
&\leqslant \Big(\mathbb{E}\sup_{t\in[0,T]} |\psi_t|^\gamma\Big)^{1/\gamma}(\mathbb{E}|\gamma_y(y^u(0)) - \gamma_y(y^v(0))|^\beta)^{1/\beta}\\
&\leqslant K\Big(\mathbb{E}\int_0^T |\tilde{k}_s|^\beta \,\mathrm{d}s\Big)^{1/\gamma}(\mathbb{E}|y^u(0) - y^v(0)|^\beta)^{1/\beta}\\
&\leqslant C\Big(\mathbb{E}\int_0^T |\tilde{k}_s|^\beta \,\mathrm{d}s\Big)^{1/\gamma} d(u(\cdot),v(\cdot))^{\alpha/2}.
\end{aligned} \tag{3.33}$$

The second term is estimated as follows. Applying the Hölder inequality yields that

$$\begin{aligned}
&\mathbb{E}\int_0^T\int_\Theta |\psi_t^* G_y(t,\theta)|\mu(\mathrm{d}\theta)\,\mathrm{d}t\\
&\leqslant \mathbb{E}\Big[\sup_{t\in[0,T]} |\psi_t| \int_0^T\int_\Theta |G_y(t,\theta)|\mu(\mathrm{d}\theta)\,\mathrm{d}t\Big]\\
&\leqslant (\mathbb{E}\sup_{t\in[0,T]} |\psi_t|^\gamma)^{1/\gamma}\Big(\mathbb{E}\int_0^T\int_\Theta |G_y(t,\theta)|^\beta\mu(\mathrm{d}\theta)\,\mathrm{d}t\Big)^{1/\beta}\\
&\leqslant K\Big(\mathbb{E}\int_0^T |\tilde{k}_s|^\beta \,\mathrm{d}s\Big)^{1/\gamma}\Big(\mathbb{E}\int_0^T\int_\Theta |G_y(t,\theta)|^\beta\mu(\mathrm{d}\theta)\,\mathrm{d}t\Big)^{1/\beta}.
\end{aligned} \tag{3.34}$$

Since $1 - 1/\gamma = 1/\beta$, one obtains putting the previous results together that

$$\mathbb{E}\int_0^T |\tilde{k}_t|^\beta \,\mathrm{d}t \leqslant Cd(u(\cdot), v(\cdot))^{\alpha\beta/2} + C\mathbb{E}\int_0^T\int_\Theta |G_y(t,\theta)|^\beta \mu(\mathrm{d}\theta)\,\mathrm{d}t. \tag{3.35}$$

Therefore, it remains to estimate the last term. Recalling that

$$\begin{aligned} G_y(t,\theta) = [g_y(t, \Lambda_t^u(\theta), u(t)) - g_y(t, \Lambda_t^v(\theta), v(t))]k_t^v \\ + l_y(t, \Lambda_t^u(\theta), u(t)) - l_y(t, \Lambda_t^v(\theta), v(t)), \end{aligned}$$

and denoting

$$\begin{aligned} &\mathbb{E}\int_0^T\int_\Theta |G_y(t,\theta)|^\beta \mu(\mathrm{d}\theta)\,\mathrm{d}t \\ &\quad\leqslant C\mathbb{E}\int_0^T\int_\Theta (|g_y(t, \Lambda_t^u(\theta), u(t)) - g_y(t, \Lambda_t^v(\theta), v(t))|^\beta |k_t^v|^\beta \\ &\qquad + |l_y(t, \Lambda_t^u(\theta), u(t)) - l_y(t, \Lambda_t^v(\theta), v(t))|^\beta)\mu(\mathrm{d}\theta)\,\mathrm{d}t \\ &\quad= \mathbb{I}_1 + \mathbb{I}_2, \end{aligned} \tag{3.36}$$

we estimate the two integrals separately. Starting with $\mathbb{I}_2$ we note that

$$\begin{aligned} \mathbb{I}_2 &\leqslant \mathbb{E}\int_0^T\int_\Theta |l_y(t, \Lambda_t^u(\theta), u(t)) - l_y(t, \Lambda_t^v(\theta), u(t))|^\beta \mu(\mathrm{d}\theta)\,\mathrm{d}t \\ &\quad + \mathbb{E}\int_0^T\int_\Theta |l_y(t, \Lambda_t^v(\theta), u(t)) - l_y(t, \Lambda_t^v(\theta), u(t))|^\beta \mathbf{1}_{\{u(t)\neq v(t)\}}(t)\mu(\mathrm{d}\theta)\,\mathrm{d}t \\ &= \mathbb{I}_2^1 + \mathbb{I}_2^2. \end{aligned} \tag{3.37}$$

Due to the growth condition on l_y in (H2) ($|l_y(t,x,y,z,r,u)| \leqslant C(1+|x|+|y|+|z|+|r|)$), the Hölder inequality with $s = 2/\beta > 1$, $r = 2/(2-\beta) > 1$ and by Lemma 3 it follows that

$$\begin{aligned} \mathbb{I}_2^2 &\leqslant C\mathbb{E}\int_0^T\int_\Theta (1 + \sup_{0\leqslant t\leqslant T}[|x^v(t)|^\beta + |y^v(t)|^\beta] + |z^v(t)|^\beta + |r_t^v(\theta)|^\beta) \\ &\quad \times \mathbf{1}_{\{u(t)\neq v(t)\}}(t)\mu(\mathrm{d}\theta)\,\mathrm{d}t \\ &\leqslant C\biggl(1 + \mathbb{E}\Bigl[\sup_{0\leqslant t\leqslant T}|x^v(t)|^2\Bigr] + \mathbb{E}\Bigl[\sup_{0\leqslant t\leqslant T}|y^v(t)|^2\Bigr] + \mathbb{E}\int_0^T |z^v(t)|^2\,\mathrm{d}t \\ &\quad + \int_0^T\int_\Theta |r_t^v(\theta)|^2\mu(\mathrm{d}\theta)\,\mathrm{d}t\biggr)^{\beta/2}\biggl(\mathbb{E}\int_0^T\int_\Theta \mathbf{1}_{\{u(t)\neq v(t)\}}(t)\mu(\mathrm{d}\theta)\,\mathrm{d}t\biggr)^{(2-\beta)/2} \\ &\leqslant Cd(u(\cdot), v(\cdot))^{\alpha\beta/2}. \end{aligned} \tag{3.38}$$

Now by the Lipschitz property of l_y and by Lemma 3 it follows that

$$\mathbb{I}_2^1 \leqslant C\mathbb{E}\int_0^T\int_\Theta (|\tilde{x}_t|^\beta + |\tilde{y}_t|^\beta + |\tilde{z}_t|^\beta + |\tilde{r}_t(\theta)|^\beta)\mu(\mathrm{d}\theta)\,\mathrm{d}t \leqslant Cd(u(\cdot),v(\cdot))^{\alpha\beta/2}.$$

Similarly we can write for $\mathbb{I}_1$

$$\begin{aligned}\mathbb{I}_1 \leqslant \mathbb{E}\int_0^T\int_\Theta &|g_y(t,\Lambda_t^u(\theta),u(t)) - g_y(t,\Lambda_t^v(\theta),u(t))|^\beta |k_t^v|^\beta \mu(\mathrm{d}\theta)\,\mathrm{d}t\\ + \mathbb{E}\int_0^T\int_\Theta &|g_y(t,\Lambda_t^v(\theta),u(t)) - g_y(t,\Lambda_t^v(\theta),v(t))|^\beta |k_t^v|^\beta \mathbf{1}_{\{u(t)\neq v(t)\}}(t)\mu(\mathrm{d}\theta)\,\mathrm{d}t.\end{aligned}$$

Due to the boundedness of g_y and its independence of y, z, r we obtain (along similar lines) the final estimate

$$\mathbb{I}_1 \leqslant Cd(u(\cdot),v(\cdot))^{\alpha\beta/2}.$$

Following similar lines we finally prove that

$$\mathbb{E}\int_0^T |\tilde{k}_t|^\beta\,\mathrm{d}t \leqslant Cd(u(\cdot),v(\cdot))^{\alpha\beta/2}.$$

Now, let us prove inequality (3.11). It is not difficult to see that $(\tilde{p}_t, \tilde{q}_t, \tilde{k}_t, \widetilde{R}_t(\theta))_{t\in[0,T]}$ satisfies the backward stochastic differential equation

$$\begin{aligned}-\mathrm{d}\tilde{p}_t = &\bigg\{ f_x^*(t,x^u(t),u(t))\tilde{p}_t + \sigma_x^*(t,x^u(t),u(t))\tilde{q}_t\\ &+ \int_\Theta [c_x^*(t,x^u(t),u(t),\theta)\tilde{R}_t(\theta) - g_x^*(t,\Lambda_t^u(\theta),u(t))\tilde{k}_t\\ &+ l_x^*(t,\Lambda_t^u(\theta),u(t)) - l_x^*(t,\Lambda_t^v(\theta),v(t))]\mu(\mathrm{d}\theta) + \mathcal{L}(t)\bigg\}\,\mathrm{d}t\\ &- \tilde{q}_t\,\mathrm{d}W(t) - \int_\Theta \widetilde{R}_t(\theta)N(\mathrm{d}\theta,\mathrm{d}t),\\ \tilde{p}_T = &-[\varphi_x^*(x^u(T))k_T^u - \varphi_x^*(x^v(T))k_T^v] + [h_x(x^u(T)) - h_x(x^v(T))],\end{aligned}$$

where the process $(\mathcal{L}(t))_{t\in[0,T]}$ is given by

$$\begin{aligned}\mathcal{L}(t) = &[f_x^*(t,x^u(t),u(t)) - f_x^*(t,x^v(t),v(t))]p_t^v\\ &+ [\sigma_x^*(t,x^u(t),u(t)) - \sigma_x^*(t,x^v(t),v(t))]q_t^v\\ &+ \int_\Theta [(c_x^*(t,x^u(t),u(t),\theta) - c_x^*(t,x^v(t),v(t),\theta))R_t^v(\theta)\\ &- (g_x^*(t,\Lambda_t^u(\theta),u(t)) - g_x^*(t,\Lambda_t^v(\theta),v(t)))k_t^v]\mu(\mathrm{d}\theta).\end{aligned}$$

Let $(\varphi_t)_{t\in[0,T]}$ be the solution of the (forward) linear SDE

$$(3.39)\quad \begin{cases} \mathrm{d}\varphi_t = [f_x^*(t,x^u(t),u(t))\varphi_t + |\tilde{p}_t|^{\beta-1}\operatorname{sgn}(\tilde{p}_t)]\,\mathrm{d}t \\ \qquad + [\sigma_x^*(t,x^u(t),u(t))\varphi_t + |\tilde{q}_t|^{\beta-1}\operatorname{sgn}(\tilde{q}_t)]\,\mathrm{d}W(t) \\ \qquad + \int_\Theta [c_x^*(t,x^u(t_-),u(t),\theta)\varphi_t + |\widetilde{R}_t(\theta)|^{\beta-1}\operatorname{sgn}(\widetilde{R}_t(\theta))]N(\mathrm{d}\theta,\mathrm{d}t), \\ \varphi_0 = 0, \end{cases}$$

where $\operatorname{sgn}(a) = (\operatorname{sgn}(a_1),\operatorname{sgn}(a_2),\ldots,\operatorname{sgn}(a_n))^*$ for any vector $a=(a_1,a_2,\ldots,a_n)^*$.

Note that since f_x,σ_x,g_x are bounded, c_x is bounded in some integral sense (see (H2)) and due to the fact that

$$\mathbb{E}\int_0^T \{||\tilde{p}_t|^{\beta-1}\operatorname{sgn}(\tilde{p}_t)|^2 + ||\tilde{q}_t|^{\beta-1}\operatorname{sgn}(\tilde{q}_t)|^2 + ||\tilde{k}_t|^{\beta-1}\operatorname{sgn}(\tilde{k}_t)|^2\}\,\mathrm{d}t \\ + \mathbb{E}\int_0^T\int_\Theta ||\widetilde{R}_t(\theta)|^{\beta-1}\operatorname{sgn}(\widetilde{R}_t(\theta))|^2\mu(\mathrm{d}\theta)\,\mathrm{d}t < \infty,$$

the linear SDE (3.39) has a unique strong solution.

Let $\gamma > 2$ be such that $1/\gamma + 1/\beta = 1$, $\beta\in(1,2)$. Then according to (2.4) we get

$$(3.40)\quad \mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}|\varphi_t|^\gamma\Big) \\ \leqslant C\mathbb{E}\int_0^T \Big\{|\tilde{p}_t|^{\beta\gamma-\gamma} + |\tilde{q}_t|^{\beta\gamma-\gamma} + |\tilde{k}_t|^{\beta\gamma-\gamma} + \int_\Theta |\widetilde{R}_t(\theta)|^{\beta\gamma-\gamma}\mu(\mathrm{d}\theta)\Big\}\,\mathrm{d}t \\ = C\mathbb{E}\int_0^T \Big\{|\tilde{p}_t|^{\beta} + |\tilde{q}_t|^{\beta} + |\tilde{k}_t|^{\beta} + \int_\Theta |\widetilde{R}_t(\theta)|^{\beta}\mu(\mathrm{d}\theta)\Big\}\,\mathrm{d}t.$$

Note that the right-hand side of the above inequality is bounded due to (2.4). Hence, we have

$$\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}|\varphi_t|^\gamma\Big) < \infty.$$

Now, by applying Itô formula to $\varphi_t^*\tilde{p}_t$ on $[0,T]$ and taking expectation, we obtain

$$\mathbb{E}[\tilde{p}_T^*\varphi_T] = \mathbb{E}\int_0^T \Big\{|\tilde{p}_t|^\beta + |\tilde{q}_t|^\beta + \int_\Theta |\widetilde{R}_t(\theta)|^\beta\mu(\mathrm{d}\theta)\Big\}\,\mathrm{d}t \\ -\mathbb{E}\int_0^T \varphi_t^*\Big(\int_\Theta [l_x(t,\Lambda_t^u(\theta),u(t)) - l_x(t,\Lambda_t^v(\theta),v(t))]\mu(\mathrm{d}\theta)\Big)\,\mathrm{d}t \\ -\mathbb{E}\int_0^T \varphi_t^*\mathcal{L}(t)\,\mathrm{d}t + \mathbb{E}\int_0^T \varphi_t^*\Big(\int_\Theta g_x^*(t,\Lambda_t^u(\theta),u(t))\tilde{k}_t\mu(\mathrm{d}\theta)\Big)\,\mathrm{d}t.$$

Now taking into account the value of $\tilde{p}_T^*$ and rearranging the terms, we deduce that

$$\mathbb{E}\int_0^T \left\{|\tilde{p}_t|^\beta + |\tilde{q}_t|^\beta + \int_\Theta |\widetilde{R}_t(\theta)|^\beta \mu(\mathrm{d}\theta)\right\} \mathrm{d}t = \mathbb{I}_1 + \mathbb{I}_2 + \mathbb{I}_3 + \mathbb{I}_4 + \mathbb{I}_5, \tag{3.41}$$

where

$$\begin{aligned}
\mathbb{I}_1 &= -\mathbb{E}(\varphi_T^*[\varphi_x^*(x^u(T))k_T^u - \varphi_x^*(x^v(T))k_T^v]),\\
\mathbb{I}_2 &= \mathbb{E}(\varphi_T^*[h_x(x^u(T)) - h_x(x^v(T))]),\\
\mathbb{I}_3 &= \mathbb{E}\int_0^T \varphi_t^* \mathcal{L}(t)\,\mathrm{d}t,\\
\mathbb{I}_4 &= \mathbb{E}\int_0^T \varphi_t^*\left(\int_\Theta [l_x^*(t,\Lambda_t^u(\theta),u(t)) - l_x^*(t,\Lambda_t^v(\theta),v(t))]\mu(\mathrm{d}\theta)\right)\mathrm{d}t,\\
\mathbb{I}_5 &= \mathbb{E}\int_0^T \varphi_t^*\left(\int_\Theta g_x^*(t,\Lambda_t^u(\theta),u(t))\tilde{k}_t\mu(\mathrm{d}\theta)\right)\mathrm{d}t.
\end{aligned}$$

Due to the Hölder inequality with $\gamma > 2$, $1/\gamma + 1/\beta = 1$, we have

$$\mathbb{I}_i \leqslant \left[\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}|\varphi_t|^\gamma\Big)\right]^{1/\gamma}\tilde{\mathbb{I}}_i \quad \text{for each } i = 1,\dots,5.$$

Then we deduce from (3.40) and (3.41) that

$$\begin{aligned}
&\mathbb{E}\int_0^T \left\{|\tilde{p}_t|^\beta + |\tilde{q}_t|^\beta + \int_\Theta |\widetilde{R}_t(\theta)|^\beta\mu(\mathrm{d}\theta)\right\}\mathrm{d}t\\
&\qquad\leqslant C\left(\mathbb{E}\int_0^T \{|\tilde{p}_t|^\beta + |\tilde{q}_t|^\beta + \int_\Theta |\widetilde{R}_t(\theta)|^\beta\mu(\mathrm{d}\theta)\}\,\mathrm{d}t\right)^{1/\gamma}\sum_{i=1}^5 \tilde{\mathbb{I}}_i,
\end{aligned}$$

and therefore (denoting $\mathbb{J}_i = \tilde{\mathbb{I}}_i^\beta$)

$$\mathbb{E}\int_0^T \left\{|\tilde{p}_t|^\beta + |\tilde{q}_t|^\beta + \int_\Theta |\widetilde{R}_t(\theta)|^\beta\mu(\mathrm{d}\theta)\right\}\mathrm{d}t \leqslant C\sum_{i=1}^5 \mathbb{J}_i.$$

Now, it remains to find appropriate estimates of $\mathbb{J}_i$, $i = 1,\dots,5$. Using (H3), Lemma 3, (3.23) and the fact that $[\mathbb{E}|k_T|^2]^{\beta/2} < \infty$, we obtain

$$\begin{aligned}
\mathbb{J}_1 &= \mathbb{E}|\varphi_x^*(x^u(T))k_T^u - \varphi_x^*(x^v(T))k_T^v|^\beta\\
&= \mathbb{E}|[\varphi_x^*(x^u(T)) - \varphi_x^*(x^v(T))]k_T^u + \varphi_x^*(x^v(T))\tilde{k}_T|^\beta\\
&\leqslant C([\mathbb{E}|k_T^u|^2]^{\beta/2}[\mathbb{E}|\varphi_x^*(x^u(T)) - \varphi_x^*(x^v(T))|^{2\beta/(2-\beta)}]^{1-\beta/2} + \mathbb{E}|\tilde{k}_T|^\beta)\\
&\leqslant C([\mathbb{E}|\tilde{x}(T)|^{2\beta/(2-\beta)}]^{1-\beta/2} + \mathbb{E}|\tilde{k}_T|^\beta) \leqslant Cd(u(\cdot),v(\cdot))^{\alpha\beta/2}.
\end{aligned} \tag{3.42}$$

Using similar arguments, we see that

$$(3.43)\qquad \mathbb{J}_2=\mathbb{E}|h_x(x^u(T))-h_x(x^v(T))|^\beta\leqslant C\mathbb{E}|\tilde{x}(T)|^\beta\leqslant Cd(u(\cdot),v(\cdot))^{\alpha\beta/2}.$$

Note that the estimation of $\mathbb{J}_4=\mathbb{E}\int_0^T\int_\Theta|l_x^*(t,\Lambda_t^u(\theta),u(t))-l_x^*(t,\Lambda_t^v(\theta),v(t))|^\beta\mu(\mathrm{d}\theta)\,\mathrm{d}t$ and of $\mathbb{J}_3=\mathbb{E}\int_0^T|\mathcal{L}(t)|^\beta\,\mathrm{d}t$ follows lines similar to the estimation of $\mathbb{I}_2$ in (3.36).

The estimate of $\mathbb{J}_5=\mathbb{E}\int_0^T\int_\Theta|g_x^*(t,\Lambda_t^u(\theta),u(t))\tilde{k}_t|^\beta\mu(\mathrm{d}\theta)\,\mathrm{d}t$ follows immediately by the boundedness of g_x and (3.23).

Finally, the desired result (3.22) follows immediately by putting all the previous estimates together. This completes the proof of Lemma 4. □

Proof of Theorem 1. First, for each $\varepsilon>0$ and $\lambda\in[0,1/2)$, by using Lemma 2 there exists $\bar{u}^\varepsilon(\cdot)\in\mathcal{U}_{\rm ad}$ and an $(\mathcal{F}_t)$-adapted process $\overline{\Psi}_t^\varepsilon(\cdot)=(\bar{p}_t^\varepsilon,\bar{q}_t^\varepsilon,\overline{k}_t^\varepsilon,\overline{R}_t^\varepsilon(\cdot))_{t\in[0,T]}$ such that for all $u\in\mathbb{U}$ we have

$$\mathbb{E}\int_0^T H_u^*(t,\overline{\Lambda}_t^\varepsilon(\cdot),\bar{u}^\varepsilon(t),\overline{\Psi}_t^\varepsilon(\cdot))(u-\bar{u}^\varepsilon(t))\,\mathrm{d}t\geqslant -C\varepsilon^\lambda.$$

Now, to prove (3.8) it remains to estimate the difference

$$\begin{aligned}\Delta^\varepsilon=\mathbb{E}&\int_0^T H_u^*(t,\overline{\Lambda}_t^\varepsilon(\cdot),\bar{u}^\varepsilon(t),\overline{\Psi}_t^\varepsilon(\cdot))(u-\bar{u}^\varepsilon(t))\,\mathrm{d}t\\&-\mathbb{E}\int_0^T H_u^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))(u-u^\varepsilon(t))\,\mathrm{d}t.\end{aligned}$$

First, by adding and subtracting the term $H_u(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))\bar{u}^\varepsilon(t)$, we have

$$\begin{aligned}\Delta^\varepsilon&=\mathbb{E}\int_0^T[H_u^*(t,\overline{\Lambda}_t^\varepsilon(\cdot),\bar{u}^\varepsilon(t),\overline{\Psi}_t^\varepsilon(\cdot))-H_u^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))](u-\bar{u}^\varepsilon(t))\,\mathrm{d}t\\&\quad+\mathbb{E}\int_0^T H_u^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))(u^\varepsilon(t)-\bar{u}^\varepsilon(t))\,\mathrm{d}t\\&=\mathbb{I}_1^\varepsilon+\mathbb{I}_2^\varepsilon.\end{aligned}$$

Now, by using the Cauchy-Schwarz inequality, boundedness of H_u in some integral sense, the fact that $\mathbb{E}\int_0^T|(u^\varepsilon(t)-\bar{u}^\varepsilon(t))|^2\,\mathrm{d}t\leqslant C$ and (3.4) we obtain for $1/\alpha+1/\gamma=1$

$$\begin{aligned}(3.44)\quad \mathbb{I}_2^\varepsilon&=\mathbb{E}\int_0^T H_u^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))(u^\varepsilon(t)-\bar{u}^\varepsilon(t))\,\mathrm{d}t\\&\leqslant\mathbb{E}\int_0^T|H_u(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))||(u^\varepsilon(t)-\bar{u}^\varepsilon(t))|\,\mathrm{d}t\\&\leqslant\left[\mathbb{E}\int_0^T|H_u(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))|^2\,\mathrm{d}t\right]^{1/2}\\&\quad\times\left[\mathbb{E}\int_0^T|(u^\varepsilon(t)-\bar{u}^\varepsilon(t))|^2\mathbf{1}_{\{u^\varepsilon(t)\neq\bar{u}^\varepsilon(t)\}}(t)\,\mathrm{d}t\right]^{1/2}\end{aligned}$$

$$
\begin{aligned}
&\leqslant C\bigg[\mathbb{E}\int_0^T |(u^\varepsilon(t)-\bar{u}^\varepsilon(t))|^{2\gamma}\,\mathrm{d}t\bigg]^{1/(2\gamma)}\bigg[\mathbb{E}\int_0^T \mathbf{1}_{\{u^\varepsilon(t)\neq \bar{u}^\varepsilon(t)\}}(t)\,\mathrm{d}t\bigg]^{\alpha/2}\\
&\leqslant Cd(u^\varepsilon(\cdot),\bar{u}^\varepsilon(\cdot))^{\alpha/2}\leqslant C\varepsilon^\lambda.
\end{aligned}
$$

Let us turn to the first term. We have

$$
\begin{aligned}
\mathbb{I}_1^\varepsilon &= \mathbb{E}\int_0^T [H_u^*(t,\overline{\Lambda}_t^\varepsilon(\cdot),\bar{u}^\varepsilon(t),\overline{\Psi}_t^\varepsilon(\cdot)) - H_u^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))](u-u^\varepsilon(t))\,\mathrm{d}t\\
&= \mathbb{E}\int_0^T [(\bar{p}_t^\varepsilon)^* f_u(t,\bar{x}^\varepsilon(t),\bar{u}^\varepsilon(t)) - (p_t^\varepsilon)^* f_u(t,x^\varepsilon(t),u^\varepsilon(t))](u-u^\varepsilon(t))\,\mathrm{d}t\\
&\quad + \mathbb{E}\int_0^T [(\bar{q}_t^\varepsilon)^*\sigma_u(t,\bar{x}^\varepsilon(t),\bar{u}^\varepsilon(t)) - (q_t^\varepsilon)^*\sigma_u(t,x^\varepsilon(t),u^\varepsilon(t))](u-u^\varepsilon(t))\,\mathrm{d}t\\
&\quad + \mathbb{E}\int_0^T\int_\Theta [(\overline{R}_t^\varepsilon(\theta))^* c_u(t,\bar{x}^\varepsilon(t),\bar{u}^\varepsilon(t),\theta)\\
&\quad - (R_t^\varepsilon(\theta))^* c_u(t,x^\varepsilon(t),u^\varepsilon(t),\theta)](u-u^\varepsilon(t))\mu(\mathrm{d}\theta)\,\mathrm{d}t\\
&\quad + \mathbb{E}\int_0^T\int_\Theta [(\bar{k}_t^\varepsilon)^* g_u(t,\overline{\Lambda}_t^\varepsilon(\theta),\bar{u}^\varepsilon(t)) - (k_t^\varepsilon)^* g_u(t,\Lambda_t^\varepsilon(\theta),u^\varepsilon(t))](u-u^\varepsilon(t))\mu(\mathrm{d}\theta)\,\mathrm{d}t\\
&\quad + \mathbb{E}\int_0^T\int_\Theta [l_u^*(t,\overline{\Lambda}_t^\varepsilon(\theta),\bar{u}^\varepsilon(t)) - l_u^*(t,\Lambda_t^\varepsilon(\theta),u^\varepsilon(t))](u-u^\varepsilon(t))\mu(\mathrm{d}\theta)\,\mathrm{d}t\\
&= \mathbb{J}_1^\varepsilon + \mathbb{J}_2^\varepsilon + \mathbb{J}_3^\varepsilon + \mathbb{J}_4^\varepsilon + \mathbb{J}_5^\varepsilon.
\end{aligned}
$$

We estimate the first term on the right-hand side $\mathbb{J}_1^\varepsilon$ by adding and subtracting $(p_t^\varepsilon)^* f_u(t,\bar{x}^\varepsilon(t),\bar{u}^\varepsilon(t))$. Then we obtain

$$
\begin{aligned}
\mathbb{J}_1^\varepsilon \leqslant{}& \mathbb{E}\int_0^T |\bar{p}_t^\varepsilon - p_t^\varepsilon||f_u(t,\bar{x}^\varepsilon(t),\bar{u}^\varepsilon(t))(u-u^\varepsilon(t))|\,\mathrm{d}t\\
&+ \mathbb{E}\int_0^T |(p_t^\varepsilon)^*(f_u(t,\bar{x}^\varepsilon(t),\bar{u}^\varepsilon(t)) - f_u(t,x^\varepsilon(t),u^\varepsilon(t)))(u-u^\varepsilon(t))|\,\mathrm{d}t.
\end{aligned}
$$

Now, again by adding and subtracting $f_u(t,x^\varepsilon(t),\bar{u}^\varepsilon(t))$, we arrive at

$$
\begin{aligned}
\mathbb{J}_1^\varepsilon &\leqslant \mathbb{E}\int_0^T |\bar{p}_t^\varepsilon - p_t^\varepsilon||f_u(t,\bar{x}^\varepsilon(t),\bar{u}^\varepsilon(t))(u-u^\varepsilon(t))|\,\mathrm{d}t\\
&\quad + \mathbb{E}\int_0^T |f_u(t,\bar{x}^\varepsilon(t),\bar{u}^\varepsilon(t)) - f_u(t,x^\varepsilon(t),\bar{u}^\varepsilon(t))||p_t^\varepsilon||u-u^\varepsilon(t)|\,\mathrm{d}t\\
&\quad + \mathbb{E}\int_0^T |f_u(t,x^\varepsilon(t),\bar{u}^\varepsilon(t)) - f_u(t,x^\varepsilon(t),u^\varepsilon(t))|\mathbf{1}_{\{\bar{u}^\varepsilon(t)\neq u^\varepsilon(t)\}}(t)|p_t^\varepsilon||u-u^\varepsilon(t)|\,\mathrm{d}t\\
&= \mathbb{J}_1^{\varepsilon,1} + \mathbb{J}_1^{\varepsilon,2} + \mathbb{J}_1^{\varepsilon,3}.
\end{aligned}
$$

Using the Hölder inequality with $1/\gamma + 1/\beta = 1$, $\beta \in (1,2)$, the boundedness of f_u and Lemma 4 leads to

$$\begin{aligned}
\mathbb{J}_1^{\varepsilon,1} &\leqslant \mathbb{E}\int_0^T |f_u(t,\bar{x}_t^\varepsilon,\bar{u}_t^\varepsilon)||u-u_t^\varepsilon||\bar{p}_t^\varepsilon - p_t^\varepsilon|\mathbf{1}_{\{u_t^\varepsilon\neq\bar{u}_t^\varepsilon\}}(t)\,\mathrm{d}t \\
&\leqslant C\biggl(\mathbb{E}\int_0^T |\bar{p}_t^\varepsilon - p_t^\varepsilon|^\beta\,\mathrm{d}t\biggr)^{1/\beta}\biggl(\mathbb{E}\int_0^T \mathbf{1}_{\{u_t^\varepsilon\neq\bar{u}_t^\varepsilon\}}(t)\,\mathrm{d}t\biggr)^{1/\gamma} \\
&\leqslant Cd^{\alpha/2}(\bar{u}^\varepsilon(\cdot),u^\varepsilon(\cdot))d^{1/\gamma}(\bar{u}^\varepsilon(\cdot),u^\varepsilon(\cdot)) \leqslant C\varepsilon^{\alpha/4+1/(2\gamma)} \leqslant C\varepsilon^\lambda;
\end{aligned}$$

since $\alpha/4 + 1/(2\gamma) \in (0,1/2)$, one can choose $\alpha/4 + 1/(2\gamma) \leqslant \lambda$.

To estimate the second term $\mathbb{J}_1^{\varepsilon,2}$ we employ the Lipschitz property of f_u from (H3) and the Hölder inequality with $1/\gamma + 1/\beta = 1$, $\beta \in (1,2)$ to obtain

$$\begin{aligned}
\mathbb{J}_1^{\varepsilon,2} &\leqslant C\mathbb{E}\int_0^T |\bar{x}_t^\varepsilon - x_t^\varepsilon||p_t^\varepsilon(u-u_t^\varepsilon)|\,\mathrm{d}t \leqslant C\mathbb{E}\int_0^T |\bar{x}_t^\varepsilon - x_t^\varepsilon||p_t^\varepsilon|\mathbf{1}_{\{u_t^\varepsilon\neq\bar{u}_t^\varepsilon\}}(t)\,\mathrm{d}t \\
&\leqslant C\biggl(\mathbb{E}\int_0^T |\bar{x}_t^\varepsilon - x_t^\varepsilon|^\gamma\,\mathrm{d}t\biggr)^{1/\gamma}\biggl(\mathbb{E}\int_0^T |p_t^\varepsilon|^\beta\mathbf{1}_{\{u_t^\varepsilon\neq\bar{u}_t^\varepsilon\}}(t)\,\mathrm{d}t\biggr)^{1/\beta} \\
&\leqslant Cd^{\alpha/2}(\bar{u}^\varepsilon(\cdot),u^\varepsilon(\cdot))\biggl(\mathbb{E}\int_0^T |p_t^\varepsilon|^2\,\mathrm{d}t\biggr)^{1/2}\biggl(\mathbb{E}\int_0^T \mathbf{1}_{\{u_t^\varepsilon\neq\bar{u}_t^\varepsilon\}}(t)\,\mathrm{d}t\biggr)^{1/\beta(1-\beta/2)} \\
&\leqslant Cd^{\alpha}(\bar{u}^\varepsilon(\cdot),u^\varepsilon(\cdot)) \leqslant C\varepsilon^{\alpha/2} \leqslant C\varepsilon^\lambda,
\end{aligned}$$

where we also used the fact that $1-\beta/2 > \alpha\beta/2$.

Next, by the Lipschitz property of f_u and by the Hölder inequality (with β and γ same as above) we can proceed to estimate $\mathbb{J}_1^3$ as follows:

$$\begin{aligned}
\mathbb{J}_1^{\varepsilon,3} &\leqslant C\mathbb{E}\int_0^T |\bar{u}_t^\varepsilon - u_t^\varepsilon||p_t^\varepsilon||u-u_t^\varepsilon|\mathbf{1}_{\{u_t^\varepsilon\neq\bar{u}_t^\varepsilon\}}(t)\,\mathrm{d}t \\
&\leqslant C\biggl(\mathbb{E}\int_0^T |p_t^\varepsilon|^\beta\mathbf{1}_{\{u_t^\varepsilon\neq\bar{u}_t^\varepsilon\}}(t)\,\mathrm{d}t\biggr)^{1/\beta}\biggl(\mathbb{E}\int_0^T |\bar{u}_t^\varepsilon - u_t^\varepsilon|^\gamma\mathbf{1}_{\{u_t^\varepsilon\neq\bar{u}_t^\varepsilon\}}(t)\,\mathrm{d}t\biggr)^{1/\gamma} \\
&\leqslant C\biggl(\mathbb{E}\int_0^T |p_t^\varepsilon|^2\,\mathrm{d}t\biggr)^{1/2}\biggl(\mathbb{E}\int_0^T \mathbf{1}_{\{u_t^\varepsilon\neq\bar{u}_t^\varepsilon\}}(t)\,\mathrm{d}t\biggr)^{1/\beta(1-\beta/2)}\biggl(\mathbb{E}\int_0^T \mathbf{1}_{\{u_t^\varepsilon\neq\bar{u}_t^\varepsilon\}}(t)\,\mathrm{d}t\biggr)^{1/\gamma} \\
&\leqslant Cd^{\alpha/2+1/\gamma}(\bar{u}^\varepsilon(\cdot),u^\varepsilon(\cdot)) \leqslant C\varepsilon^{\alpha/4+1/(2\gamma)} \leqslant C\varepsilon^\lambda.
\end{aligned}$$

Using arguments similar to those developed above for $\mathbb{J}_2^\varepsilon$, $\mathbb{J}_3^\varepsilon$, $\mathbb{J}_4^\varepsilon$, and $\mathbb{J}_5^\varepsilon$, we can prove that $\mathbb{I}_1^\varepsilon \leqslant C\varepsilon^\lambda$, and finally conclude that

$$\begin{aligned}
(3.45)\qquad &\mathbb{E}\int_0^T [H_u^*(t,\overline{\Lambda}_t^\varepsilon(\cdot),\bar{u}^\varepsilon(t),\overline{\Psi}_t^\varepsilon(\cdot)) - H_u^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))](u-u^\varepsilon(t))\,\mathrm{d}t \\
&\leqslant C\varepsilon^\lambda,
\end{aligned}$$

which together with (3.44) implies that

$$\Delta^{\varepsilon} \leqslant C\varepsilon^{\lambda}. \tag{3.46}$$

Now, combining (3.3) in Lemma 2 and (3.46) we complete the proof of Theorem 1. □

4. Sufficient conditions for near-optimality of FBSDEJs

In this section, we will prove that under an additional hypothesis, the near-maximality condition on the Hamiltonian function is a sufficient condition for near-optimality. This is the second main result of this paper.

Theorem 2. (Sufficient Near-Optimality Maximum Principle). *Let $u^{\varepsilon}(\cdot)$ be an admissible control and let the processes $\Lambda_t^{\varepsilon}(\cdot) = (x^{\varepsilon}(t), y^{\varepsilon}(t), z^{\varepsilon}(t), r^{\varepsilon}(t,\cdot))$ and $\Psi_t^{\varepsilon}(\cdot) = (p_t^{\varepsilon}, q_t^{\varepsilon}, k_t^{\varepsilon}, R_t^{\varepsilon}(\cdot))$ be the solutions to equation* (1.1) *and adjoint equations* (2.1)*, respectively, both associated with $u^{\varepsilon}(\cdot)$.*

Further, let us assume that the function $H(t,\cdot,\cdot,\cdot,\cdot,\Psi_t^{\varepsilon}(\cdot))$ is convex for a.e. $t \in [0,T]$, $\mathbb{P}$-a.s. and functions γ, h are convex. If for some $\lambda \in [0,1/2)$, $\varepsilon > 0$ and for any $u \in \mathbb{U}$ the near-maximality relation

$$\mathbb{E}\int_0^T H_u^*(t,\Lambda_t^{\varepsilon}(\cdot),u^{\varepsilon}(t),\Psi_t^{\varepsilon}(\cdot))(u-u^{\varepsilon}(t))\,\mathrm{d}t \geqslant -C\varepsilon^{\lambda} \tag{4.1}$$

holds then we have

$$J(u^{\varepsilon}(\cdot)) \leqslant \inf_{v(\cdot)\in\mathcal{U}_{\mathrm{ad}}} J(v(\cdot)) + C\varepsilon^{\lambda},$$

where C is a positive constant independent of ε.

In other words, the process $u^{\varepsilon}(\cdot)$ is a near-optimal control of order λ to the control problem (1.1)–(1.2).

Proof. Let us fix an arbitrary $v(\cdot) \in \mathcal{U}_{\mathrm{ad}}$ and denote its corresponding trajectory $\Lambda_t^{v}(\cdot) = (x^{v}(t), y^{v}(t), z^{v}(t), r^{v}(t,\cdot))$. Then we have

$$\begin{aligned} J(u^{\varepsilon}(\cdot)) - J(v(\cdot)) &= \mathbb{E}[h(x^{\varepsilon}(T)) - h(x^{v}(T))] + \mathbb{E}[\gamma(y^{\varepsilon}(0)) - \gamma(y^{v}(0))] \\ &\quad + \mathbb{E}\int_0^T\int_{\Theta}\{l(t,\Lambda_t^{\varepsilon}(\theta),u^{\varepsilon}(t)) - l(t,\Lambda_t^{v}(\theta),v(t))\}\mu(\mathrm{d}\theta)\,\mathrm{d}t. \end{aligned}$$

Now, since h and γ are convex, we have

$$\begin{aligned}h(x^v(T)) - h(x^\varepsilon(T)) &\geqslant h_x^*(x^\varepsilon(T))(x^v(T) - x^\varepsilon(T)),\\ \gamma(y^v(0)) - \gamma(y^\varepsilon(0)) &\geqslant \gamma_y^*(y^\varepsilon(0))(y^v(0) - y^\varepsilon(0)).\end{aligned}$$

Therefore, one easily obtains that

$$\begin{aligned}J(u^\varepsilon(\cdot)) - J(v(\cdot)) \leqslant{}& \mathbb{E}[h_x^*(x^\varepsilon(T))(x^\varepsilon(T) - x^v(T))] + \mathbb{E}[\gamma_y^*(y^\varepsilon(0))(y^\varepsilon(0) - y^v(0))]\\ &+ \mathbb{E}\int_0^T\int_\Theta \{l(t, \Lambda_t^\varepsilon(\theta), u^\varepsilon(t)) - l(t, \Lambda_t^v(\theta), v(t))\}\mu(\mathrm{d}\theta)\,\mathrm{d}t.\end{aligned}$$

Next, employing the initial and terminal conditions of the adjoint equations $\gamma_y(y^\varepsilon(0)) = k_0^\varepsilon$ and $h_x(x^\varepsilon(T)) = p_T^\varepsilon + \varphi_x^*(x^\varepsilon(T))k_T^\varepsilon$, it follows that

$$\begin{aligned}J(u^\varepsilon(\cdot)) - J(v(\cdot)) \leqslant{}& \mathbb{E}[(p_T^\varepsilon)^*(x^\varepsilon(T) - x^v(T))]\\ &+ \mathbb{E}[(\varphi_x^*(x^\varepsilon(T))k_T^\varepsilon)^*(x^\varepsilon(T) - x^v(T))] + \mathbb{E}[(k_0^\varepsilon)^*(y^v(0) - y^\varepsilon(0))]\\ &+ \mathbb{E}\int_0^T\int_\Theta \{l(t, \Lambda_t^\varepsilon(\theta), u^\varepsilon(t)) - l(t, \Lambda_t^v(\theta), v(t))\}\mu(\mathrm{d}\theta)\,\mathrm{d}t.\end{aligned} \tag{4.2}$$

On the other hand, by applying Itô formula to $(p_t^\varepsilon)^*(x^\varepsilon(t) - x^v(t))$ on $[0,T]$ and taking expectation, we get

$$\begin{aligned}&\mathbb{E}[(p_T^\varepsilon)^*(x^\varepsilon(T) - x^v(T)) + (\varphi_x^*(x^\varepsilon(T))k_T^\varepsilon)^*(x^\varepsilon(T) - x^v(T))]\\ &\quad= \mathbb{E}\int_0^T \Big\{ -f_x^*(t, x^\varepsilon(t), u^\varepsilon(t))p_t^\varepsilon - \sigma_x^*(t, x^\varepsilon(t), u^\varepsilon(t))q_t^\varepsilon\\ &\qquad+ \int_\Theta g_x^*(t, \Lambda_t^\varepsilon(\theta), u^\varepsilon(t))k_t^\varepsilon\mu(\mathrm{d}\theta)\\ &\qquad- \int_\Theta [c_x^*(t, x^\varepsilon(t), u^\varepsilon(t), \theta)R_t^\varepsilon(\theta) + l_x^*(t, \Lambda_t^\varepsilon(\theta), u^\varepsilon(t))]\mu(\mathrm{d}\theta)\\ &\qquad+ (p_t^\varepsilon)^*[f(t, x^\varepsilon(t), u^\varepsilon(t)) - f(t, x^v(t), v(t))]\\ &\qquad+ \mathrm{Tr}\{(q_t^\varepsilon)^*[\sigma(t, x^\varepsilon(t), u^\varepsilon(t)) - \sigma(t, x^v(t), v(t))]\}\\ &\qquad+ \int_\Theta (R_t^\varepsilon(\theta))^*[c(t, x^\varepsilon(t), u^\varepsilon(t), \theta) - c(t, x^v(t), v(t), \theta)]\mu(\mathrm{d}\theta)\Big\}\\ &\qquad\qquad\times (x^\varepsilon(t) - x^v(t))\,\mathrm{d}t + \mathbb{E}[h_x^*(x^\varepsilon(T))(x^\varepsilon(T) - x^v(T))].\end{aligned} \tag{4.3}$$

Similarly, by applying Itô formula to $k_t^\varepsilon(y^\varepsilon(t) - y^v(t))$ then combining (4.2), (4.3), and using the definition of the Hamiltonian function, we obtain

$$\begin{aligned}
&\mathbb{E}\int_0^T\int_\Theta \{l(t,\Lambda_t^\varepsilon(\theta),u^\varepsilon(t)) - l(t,\Lambda_t^v(\theta),v(t))\}\mu(\mathrm{d}\theta)\,\mathrm{d}t \\
&\quad = \mathbb{E}\int_0^T \{H(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot)) - H(t,\Lambda_t^v(\cdot),u^v(t),\Psi_t^v(\cdot))\}\,\mathrm{d}t \\
&\qquad + \mathbb{E}\int_0^T \{-(p_t^\varepsilon)^*[f(t,x^\varepsilon(t),u^\varepsilon(t)) - f(t,x^v(t),v(t))] \\
&\qquad - \mathrm{Tr}\{(q_t^\varepsilon)^*[\sigma(t,x^\varepsilon(t),u^\varepsilon(t)) - \sigma(t,x^v(t),v(t))]\}\}\,\mathrm{d}t \\
&\qquad + \mathbb{E}\int_0^T\int_\Theta \{(k_t^\varepsilon)^*[g(t,\Lambda_t^\varepsilon(\theta),u^\varepsilon(t)) - g(t,\Lambda_t^v(\theta),v(t))] \\
&\qquad - (R_t^\varepsilon(\theta))^*[c(t,x^\varepsilon(t),u^\varepsilon(t),\theta) - c(t,x^v(t),v(t),\theta)]\}\mu(\mathrm{d}\theta)\,\mathrm{d}t.
\end{aligned}$$

Then we have

$$\begin{aligned}
(4.4)\qquad & J(u^\varepsilon(\cdot)) - J(v(\cdot)) \\
&\leqslant \mathbb{E}\int_0^T [H(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot)) - H(t,\Lambda_t^\varepsilon(\cdot),v(t),\Psi_t^\varepsilon(\cdot))]\,\mathrm{d}t \\
&\quad - \mathbb{E}\int_0^T H_x^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon,\Psi_t^\varepsilon(\cdot))(x^\varepsilon(t) - x^v(t))\,\mathrm{d}t \\
&\quad - \mathbb{E}\int_0^T H_y^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))(y^\varepsilon(t) - y^v(t))\,\mathrm{d}t \\
&\quad - \mathbb{E}\int_0^T H_z^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))(z^\varepsilon(t) - z^v(t))\,\mathrm{d}t \\
&\quad - \mathbb{E}\int_0^T H_r^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))(r_t^\varepsilon(\cdot) - r_t^v(\cdot))\,\mathrm{d}t.
\end{aligned}$$

Since $H(t,\cdot,\cdot,\cdot,\cdot,\cdot,\Psi_t^\varepsilon(\cdot))$ is convex in (x,y,z,r,u), we obtain

$$\begin{aligned}
&H(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot)) - H(t,\Lambda_t^v(\cdot),v(t),\Psi_t^\varepsilon(\cdot)) \\
&\quad \leqslant H_x^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))(x^\varepsilon(t) - x^v(t)) \\
&\qquad + H_y^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))(y^\varepsilon(t) - y^v(t)) \\
&\qquad + H_z^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))(z^\varepsilon(t) - z^v(t)) \\
&\qquad + H_r^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))(r_t^\varepsilon(\cdot) - r_t^v(\cdot)) \\
&\qquad + H_u^*(t,\Lambda_t^\varepsilon(\cdot),u^\varepsilon(t),\Psi_t^\varepsilon(\cdot))(u^\varepsilon(t) - v(t)).
\end{aligned}$$

By integrating both sides and noting (4.1), we obtain

$$
\begin{aligned}
(4.5)\qquad & \mathbb{E}\int_0^T \{H(t,\Lambda_t^{\varepsilon}(\cdot),u^{\varepsilon}(t),\Psi_t^{\varepsilon}(\cdot)) - H(t,\Lambda_t^{v}(\cdot),v(t),\Psi_t^{\varepsilon}(\cdot))\}\,\mathrm{d}t \\
&\leqslant \mathbb{E}\int_0^T H_x^*(t,\Lambda_t^{\varepsilon}(\cdot),u^{\varepsilon}(t),\Psi_t^{\varepsilon}(\cdot))(x^{\varepsilon}(t)-x^{v}(t))\,\mathrm{d}t \\
&\quad + \mathbb{E}\int_0^T H_y^*(t,\Lambda_t^{\varepsilon}(\cdot),u^{\varepsilon}(t),\Psi_t^{\varepsilon}(\cdot))(y^{\varepsilon}(t)-y^{v}(t))\,\mathrm{d}t \\
&\quad + \mathbb{E}\int_0^T H_z^*(t,\Lambda_t^{\varepsilon}(\cdot),u^{\varepsilon}(t),\Psi_t^{\varepsilon}(\cdot))(z^{\varepsilon}(t)-z^{v}(t))\,\mathrm{d}t \\
&\quad + \mathbb{E}\int_0^T H_r^*(t,\Lambda_t^{\varepsilon}(\cdot),u^{\varepsilon}(t),\Psi_t^{\varepsilon}(\cdot))(r_t^{\varepsilon}(\cdot)-r_t^{v}(\cdot))\,\mathrm{d}t + C\varepsilon^{\lambda}.
\end{aligned}
$$

Combining (4.1) and (4.5), we get

$$J(u^{\varepsilon}(\cdot)) - J(v(\cdot)) \leqslant C\varepsilon^{\lambda}.$$

Finally, since $v(\cdot)$ is arbitrary, the desired result follows. □

5. Application to finance: Mixed problem of mean-variance portfolio selection

In this section we will apply our maximum principle of near-optimality to study a perturbed mean-variance portfolio selection problem mixed with a recursive utility functional optimization in a financial market and we will derive the explicit expression for the near-optimal portfolio selection strategy. The near-optimal control will be taken as the optimal control to the unperturbed problem (see Shi and Wu [24]) and we will show that this candidate satisfies the assumptions of Theorem 2.

Suppose that we are given a mathematical market consisting of two investment possibilities (see Framstad, Øksendal and Sulam [9]):

(1) *Bond price*: the first asset is a risk-free security whose price $P_0(t)$ evolves according to the ordinary differential equation

$$
(5.1)\qquad \begin{cases} \mathrm{d}P_0(t) = P_0(t)\varrho_t\,\mathrm{d}t, \quad t\in[0,T], \\ P_0(0) > 0, \end{cases}
$$

where $\varrho\colon [0,T]\to\mathbb{R}_+$ is a locally bounded deterministic function.

(2) *Stock price*: a risky security (e.g. a stock), where the price $P_1(t)$ at time t is given by

$$
(5.2)\qquad \begin{cases} \mathrm{d}P_1(t) = P_1(t_-)\Big[\varsigma_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}W(t) + \displaystyle\int_{\Theta}\xi_t(\theta)N(\mathrm{d}\theta,\mathrm{d}t)\Big], \\ P_1(0) > 0, \end{cases}
$$

where $\varsigma, \sigma\colon [0,T] \to \mathbb{R}$ are bounded deterministic functions such that $\varsigma_t, \sigma_t \neq 0$ and $\varsigma_t > \varrho_t$ for all $t \in [0,T]$.

(3) *Assumptions.* In order to ensure that $P_1(t) > 0$ for all $t \in [0,T]$ we assume that

(i) $\xi_t(\theta) > -1$ for μ-almost all $\theta \in \Theta$ and all $t \in [0,T]$,

(ii) $\int_\Theta \xi_t^2(\theta)\mu(\mathrm{d}\theta)$ is bounded.

(4) *Portfolio strategy.* A portfolio is an $(\mathcal{F}_t)$-predictable process $(e_0(t), e_1(t))$ giving the number of units of the risk-free and the risky security held at time t. Let $\pi_t = e_1(t)P_1(t)$ denote the amount invested in the risky security. We call the control process $\pi(\cdot)$ a portfolio strategy.

(5) *The wealth dynamics.* Let $x^\pi(0) = \zeta > 0$ be the initial wealth. By combining (5.1) and (5.2) we introduce the wealth dynamics

$$(5.3)\qquad \begin{cases} \mathrm{d}x^\pi(t) = [\varrho_t x^\pi(t) + (\varsigma_t - \varrho_t)\pi_t]\,\mathrm{d}t + \sigma_t\pi_t\,\mathrm{d}W(t) + \displaystyle\int_\Theta \xi_t(\theta)\pi_t N(\mathrm{d}\theta, \mathrm{d}t), \\ x^\pi(0) = \zeta. \end{cases}$$

Let $\mathbb{U}$ be a compact convex subset of $\mathbb{R}$. We denote by $\mathcal{U}_{\mathrm{ad}}$ the set of admissible $(\mathcal{F}_t)$-predictable portfolio strategies $\pi(\cdot)$ valued in $\mathbb{U}$.

The mean-variance portfolio selection problem in the above jump-diffusion framework has been studied in [9], [25]. In Framstad, Øksendal and Sulam [9], the investor's object is to find an admissible portfolio which minimizes the variance $\mathrm{Var}(x^\pi(T))$ at a future time $T > 0$ under the condition that $\mathbb{E}[x^\pi(T)] = a$ for some given $a \in \mathbb{R}_+$. By using sufficient maximum principle, the authors in [9] gave the expression for the optimal portfolio selection. Optimal portfolio and consumption decision problems for a small investor in a market model have been studied in [14], [15]. The near-optimal consumption-investment problem has been discussed in Hafayed, Veverka and Abbas [12]. Stochastic optimization problems with recursive utility have important economic background, see [8], [7]. The continuous time mean-variance portfolio selection problem has been studied in Zhou [34].

In this section, the objective is to use our near-optimal maximum principle to study the mean-variance portfolio selection problem mixed with a recursive utility functional maximization. We consider a small investor endowed with an initial wealth $x^\pi(0) > 0$ who chooses at each time t his or her portfolio strategy π_t. The investor wants to choose a portfolio strategy $\pi^\varepsilon(\cdot) \in \mathcal{U}_{\mathrm{ad}}$ which near-maximizes the expected utility functional. This functional can be separated into two parts: the former is the equivalent terminal reward $\mathbb{E}[-1/2(x^\pi(t) - a)^2]$ while the latter part is a recursive utility functional with generator $g(t,x,y,\pi) = \varrho_t x + (\varsigma_t - \varrho_t)\pi - cy$, $c > 0$.

We assume that we originally have a family of optimization problems parameterized by a parameter $\varepsilon > 0$ representing the complexity of the cost functional

$$J^{\varepsilon}(\pi(\cdot)) = \mathbb{E}\Big\{ \int_0^T \varepsilon\varphi(\pi_t)\,\mathrm{d}t + \frac{1}{2}(x^{\pi}(T) - a)^2 \Big\} + y(0), \tag{5.4}$$

where $\varphi\colon \mathbb{R} \to \mathbb{R}$ is a nonlinear, convex and continuously differentiable function independent of ε. Further, we define the wealth process $(x(t))$ and the recursive utility process $(y(t))$ corresponding to $\pi(\cdot) \in \mathcal{U}_{\mathrm{ad}}$ as the solutions to the FBSDEJs

$$\begin{cases} \mathrm{d}x(t) = [\varrho_t x(t) + (\varsigma_t - \varrho_t)\pi_t]\,\mathrm{d}t + \sigma_t\pi_t\,\mathrm{d}W(t) + \displaystyle\int_{\Theta} \xi_t(\theta)\pi_t N(\mathrm{d}\theta, \mathrm{d}t), \\ -\mathrm{d}y(t) = [\varrho_t x(t) + (\varsigma_t - \varrho_t)\pi_t - cy(t)]\,\mathrm{d}t - z(t)\,\mathrm{d}W(t) \\ \qquad\qquad - \displaystyle\int_{\Theta} r_t(\theta)\pi_t N(\mathrm{d}\theta, \mathrm{d}t), \\ x(0) = \zeta, \quad y(T) = x(T). \end{cases} \tag{5.5}$$

We notice that setting $\varepsilon = 0$ in (5.4) leads to

$$J^{0}(\pi(\cdot)) = \mathbb{E}\Big\{ \frac{1}{2}(x^{\pi}(T) - a)^2 \Big\} + y(0). \tag{5.6}$$

The optimal control to the problem (5.5)–(5.6) (with the new cost $J^0(\cdot)$) has already been solved explicitly by using the stochastic maximum principle in Shi and Wu [25], Theorem 3.1 where the optimal solution, denoted by $(x^{\star}, \pi^{\star})$, is given in the state feedback form as

$$\begin{cases} \pi_t^{\star} = \dfrac{(\varrho_t - \varsigma_t)(\Phi_t x^{\star}(t) + \Psi_t - \exp(-ct))}{\Phi_t(\sigma_t^2 + \int_{\Theta} \xi_t^2(\theta)\mu(\mathrm{d}\theta))}, \\ p_t^{\star} = \Phi_t\pi_t^{\star} + \Psi_t, \\ q_t^{\star} = \sigma_t\Phi_t\pi_t^{\star}, \\ k_t^{\star} = \exp(-ct), \\ R_t^{\star}(\theta) = \Phi_t\xi_t(\theta)\pi_t^{\star}, \end{cases} \tag{5.7}$$

where Φ_t and Ψ_t are some deterministic differentiable functions satisfying the ordinary differential equations

$$\begin{cases} \Phi_t' = \Big(\dfrac{(\varrho_t - \varsigma_t)^2}{\sigma_t^2 + \int_{\Theta} \xi_t^2(\theta)\mu(\mathrm{d}\theta)} - 2\varrho_t\Big)\Phi_t, \\ \Phi_T = 1, \end{cases} \tag{5.8}$$

and

$$(5.9)\qquad \begin{cases} \Psi_t' = \Big(\dfrac{(\varrho_t - \varsigma_t)^2}{\sigma_t^2 + \int_\Theta \xi_t^2(\theta)\mu(\mathrm{d}\theta)} - \varrho_t\Big)\Psi_t - \mathrm{e}^{-ct}\Big(\dfrac{(\varrho_t - \varsigma_t)^2}{\sigma_t^2 + \int_\Theta \xi_t^2(\theta)\mu(\mathrm{d}\theta)} - \varrho_t\Big), \\ \Psi_T = -(a+1). \end{cases}$$

The Hamiltonian of problem (5.5)–(5.6) is given by

$$(5.10)\qquad \begin{aligned} &H(t,x,y,z,r(\cdot),\pi,p,q,k,R(\cdot)) \\ &\qquad = [\varrho_t x + (\varsigma_t - \varrho_t)\pi](k-p) + \sigma_t q\pi + \pi\int_\Theta \xi_t(\theta)R(\theta)\mu(\mathrm{d}\theta). \end{aligned}$$

Since $\pi^\star(\cdot)$ is optimal for the problem (5.5)–(5.6), by using maximum condition ([25], Theorem 2.1) we conclude that

$$(5.11)\qquad (\varsigma_t - \varrho_t)(p_t^\star - k_t^\star) + \sigma_t q_t^\star + \int_\Theta \xi_t(\theta)R_t^\star(\theta)\mu(\mathrm{d}\theta) = 0, \quad \mathbb{P}\text{-a.s., } \mathrm{d}t\text{-a.e.},$$

where we denote by the superscript * all the processes computed for the optimal control $\pi^\star(\cdot)$.

However, the Hamiltonian H^ε for the problem (5.5)–(5.4) can be rewritten in the form

$$(5.12)\qquad \begin{aligned} &H^\varepsilon(t,x,y,z,r(\cdot),\pi,p,q,k,R(\cdot)) \\ &\quad = [\varrho_t x + (\varsigma_t - \varrho_t)\pi](k-p) + \sigma_t q\pi + \pi\int_\Theta \xi_t(\theta)R(\theta)\mu(\mathrm{d}\theta) - \varepsilon\varphi(\pi) \\ &\quad = H(t,x,y,z,r(\cdot),\pi,p,q,k,R(\cdot)) - \varepsilon\varphi(\pi) \end{aligned}$$

for all $(x,y,z,r(\cdot),\pi,p,q,k,R(\cdot))$. Therefore, if $(x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot))$ denotes the optimal trajectory to the (unperturbed) control problem (5.5)–(5.6) we can express the difference of the Hamiltonian at different control points but at this fixed optimal trajectory in the following way:

$$(5.13)\qquad \begin{aligned} &H^\varepsilon(t,x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),\pi,p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot)) \\ &\qquad - H^\varepsilon(t,x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),\pi_t^\star,p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot)) \\ &\quad = H(t,x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),\pi,p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot)) \\ &\qquad - H(t,x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),\pi_t^\star,p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot)) - \varepsilon[\varphi(\pi) - \varphi(\pi_t^\star)]. \end{aligned}$$

By virtue of the fact that the function $\varphi(\cdot)$ is continuously differentiable and $\mathbb{U}$ is a compact convex subset in $\mathbb{R}$ it follows that

$$-\varepsilon[\varphi(\pi) - \varphi(\pi_t^\star)] \leqslant \varepsilon|\varphi'(\pi)||\pi - \pi_t^\star| \leqslant C\varepsilon.$$

Now, employing the above fact, taking $\max_{\pi\in\mathbb{U}}$ in (5.13) and using the optimality of $\pi^\star$ we arrive at

$$\begin{aligned}
&\max_{\pi\in\mathbb{U}} H^\varepsilon(t,x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),\pi,p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot))\\
&\qquad - H^\varepsilon(t,x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),\pi_t^\star,p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot))\\
&\quad\leqslant \max_{\pi\in\mathbb{U}} H(t,x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),\pi,p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot))\\
&\qquad - H(t,x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),\pi_t^\star,p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot)) + \varepsilon\max_{\pi\in\mathbb{U}}\{|\varphi'(\pi)||\pi_t-\pi_t^\star|\}\\
&\quad\leqslant H(t,x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),\pi_t^\star,p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot))\\
&\qquad - H(t,x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),\pi_t^\star,p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot)) + C\varepsilon = C\varepsilon,
\end{aligned}$$

which implies the near-maximality property of $\pi^\star(\cdot)$

$$\begin{aligned}
&H^\varepsilon(t,x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),\pi_t^\star,p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot))\\
&\quad\geqslant \max_{\pi\in\mathbb{U}} H^\varepsilon(t,x^\star(t),y_t^\star,z_t^\star,r_t^\star(\cdot),\pi,p_t^\star,q_t^\star,k_t^\star,R_t^\star(\cdot)) - C\varepsilon.
\end{aligned}$$

Finally, since the function $\varphi(\cdot)$ is convex, the Hamiltonian H^ε is concave. Due to the sufficient maximum principle (Theorem 2), the portfolio strategy $\pi^\star(\cdot)$ is indeed a near-optimal for the problem (5.5)–(5.4).

Remark. When $\varepsilon\to 0$, our result reduces to the necessary and sufficient conditions of optimality developed in Shi and Wu [23].

Appendix

The following result is a special case of the Itô formula for jump diffusions.

Lemma A (Integration by parts formula for jumps processes)**.** *Suppose that the processes $x_1(t)$ and $x_2(t)$ are given for $j=1,2$, $t\in[s,T]$:*

$$\begin{cases}
\mathrm{d}x_j(t) = f(t,x_j(t),u(t))\,\mathrm{d}t + \sigma(t,x_j(t),u(t))\,\mathrm{d}W(t)\\
\qquad\qquad + \displaystyle\int_\Theta g(t,x_j(t_-),u(t),\theta)N(\mathrm{d}\theta,\mathrm{d}t),\\
x_j(s) = 0.
\end{cases}$$

Then

$$\begin{aligned}
&\mathbb{E}[x_1(T)x_2(T)]\\
&\ = \mathbb{E}\left[\int_s^T x_1(t)\,\mathrm{d}x_2(t) + \int_s^T x_2(t)\,\mathrm{d}x_1(t)\right] + \mathbb{E}\int_s^T \sigma^*(t,x_1(t),u(t))\sigma(t,x_2(t),u(t))\,\mathrm{d}t\\
&\quad + \mathbb{E}\int_s^T\int_\Theta g^*(t,x_1(t),u(t),\theta)g(t,x_2(t),u(t),\theta)\mu(\mathrm{d}\theta)\,\mathrm{d}t.
\end{aligned}$$

See Framstad, Øksendal and Sulam ([9], Lemma 2.1) for the detailed proof of the above lemma.

A c k n o w l e d g e m e n t. We are thankful to the anonymous referees and the handling editor for their expert comments and suggestions which helped us to improve the manuscript considerably.

References

[1] *K. Bahlali, N. Khelfallah, B. Mezerdi*: Necessary and sufficient conditions for near-optimality in stochastic control of FBSDEs. Syst. Control Lett. *58* (2009), 857–864. zbl MR

[2] *R. Bellman*: Dynamic Programming. With a new introduction by Stuart Dreyfus. Reprint of the 1957 edition. Princeton Landmarks in Mathematics, Princeton University Press, Princeton, 2010. zbl MR

[3] *R. Boel, P. Varaiya*: Optimal control of jump processes. SIAM J. Control Optim. *15* (1977), 92–119. zbl MR

[4] *B. Bouchard, R. Elie*: Discrete-time approximation of decoupled forward-backward SDE with jumps. Stochastic Processes Appl. *118* (2008), 53–75. zbl MR

[5] *A. Cadenillas*: A stochastic maximum principle for systems with jumps, with applications to finance. Syst. Control Lett. *47* (2002), 433–444. zbl MR

[6] *I. Ekeland*: On the variational principle. J. Math. Anal. Appl. *47* (1974), 324–353. zbl MR

[7] *N. El Karoui, S. G. Peng, M. C. Quenez*: Backward stochastic differential equations in finance. Math. Finance *7* (1997), 1–71. zbl MR

[8] *N. El Karoui, S. G. Peng, M. C. Quenez*: A dynamic maximum principle for the optimization of recursive utilities under constraints. Ann. Appl. Probab. *11* (2001), 664–693. zbl MR

[9] *N. C. Framstad, B. Øksendal, A. Sulem*: Sufficient stochastic maximum principle for the optimal control of jump diffusions and applications to finance. J. Optim. Theory. Appl. *121* (2004), 77–98; erratum ibid. *124* (2005), 511–512. zbl MR

[10] *R. Gabasov, F. M. Kirillova, B. Sh. Mordukhovich*: The ε-maximum principle for suboptimal controls. Sov. Math., Dokl. *27* (1983), 95–99; translation from Dokl. Akad. Nauk SSSR *268* (1983), 525–529. (In Russian.) zbl MR

[11] *M. Hafayed, S. Abbas, P. Veverka*: On necessary and sufficient conditions for near-optimal singular stochastic controls. Optim. Lett. *7* (2013), 949–966. zbl MR

[12] *M. Hafayed, P. Veverka, S. Abbas*: On maximum principle of near-optimality for diffusions with jumps, with application to consumption-investment problem. Differ. Equ. Dyn. Syst. *20* (2012), 111–125. zbl MR

[13] *J. Huang, X. Li, G. Wang*: Near-optimal control problems for linear forward-backward stochastic systems. Automatica *46* (2010), 397–404. zbl MR

[14] *M. Jeanblanc-Picqué, M. Pontier*: Optimal portfolio for a small investor in a market model with discontinuous prices. Appl. Math. Optimization *22* (1990), 287–310. zbl MR

[15] *I. Karatzas, J. P. Lehoczky, S. E. Shreve*: Optimal portfolio and consumption decisions for a "Small investor" on a finite horizon. SIAM J. Control Optim. *25* (1987), 1557–1586. zbl MR

[16] *B. Sh. Mordukhovich*: Approximation Methods in Problems of Optimization and Control. Nauka, Moskva, 1988. (In Russian.) zbl MR

[17] *B. Øksendal, A. Sulem*: Applied Stochastic Control of Jump Diffusions. Second edition. Universitext, Springer, Berlin, 2007. zbl MR

[18] *L. P. Pan, K. L. Teo*: Near-optimal controls of class of Volterra integral systems. J. Optimization Theory Appl. *101* (1999), 355–373. zbl MR

[19] *S. Peng, Z. Wu*: Fully coupled forward-backward stochastic differential equations and application to optimal control. SIAM J. Control Optim. *37* (1999), 825–843. zbl MR
[20] *L. S. Pontryagin, V. G. Boltanskii, R. V. Gamkrelidze, E. F. Mishchenko*: The Mathematical Theory of Optimal Processes. Translation from the Russian, Interscience Publishers, New York, 1962. zbl MR
[21] *R. Rishel*: A minimum principle for controlled jump processes. Control Theory, Numer. Meth., Computer Syst. Mod.; Internat. Symp. Rocquencourt 1974, Lecture Notes Econ. Math. Syst. *107* (1975), 493–508. zbl MR
[22] *J. Shi*: Necessary conditions for optimal control of forward-backward stochastic systems with random jumps. Int. J. Stoch. Anal. 2012 (2012), Article ID 258674, 50 pp. zbl MR
[23] *J. Shi, Z. Wu*: The maximum principle for fully coupled forward-backward stochastic control system. Acta Autom. Sin. *32* (2006), 161–169. MR
[24] *J. Shi, Z. Wu*: Maximum principle for fully coupled forward-backward stochastic control system with random jumps. Proceedings of the 26th Chinese Control Conference, Zhangjiajie, Hunan, 2007, pp. 375–380.
[25] *J. Shi, Z. Wu*: Maximum principle for forward-backward stochastic control system with random jumps and applications to finance. J. Syst. Sci. Complex. *23* (2010), 219–231. zbl MR
[26] *R. Situ*: A maximum principle for optimal controls of stochastic systems with random jumps. Proceedings of National Conference on Control Theory and its Applications. Qingdao, China, 1991.
[27] *S. L. Tang, X. J. Li*: Necessary conditions for optimal control of stochastic systems with random jumps. SIAM J. Control Optim. *32* (1994), 1447–1475. zbl MR
[28] *W. Xu*: Stochastic maximum principle for optimal control problem of forward and backward system. J. Aus. Math. Soc., Ser. B *37* (1995), 172–185. zbl MR
[29] *J. Yong*: Optimality variational principle for controlled forward-backward stochastic differential equations with mixed intial-terminal conditions. SIAM J. Control. Optim. *48* (2010), 4119–4156. zbl MR
[30] *J. Yong, X. Y. Zhou*: Stochastic Controls. Hamiltonian Systems and HJB Equations. Applications of Mathematics 43, Springer, New York, 1999. zbl MR
[31] *X. Y. Zhou*: Deterministic near-optimal control. I: Necessary and sufficient conditions for near-optimality. J. Optimization Theory Appl. *85* (1995), 473–488. zbl MR
[32] *X. Y. Zhou*: Deterministic near-optimal controls. II: Dynamic programming and viscosity solution approach. Math. Oper. Res. *21* (1996), 655–674. zbl MR
[33] *X. Y. Zhou*: Stochastic near-optimal controls: Necessary and sufficient conditions for near-optimality. SIAM J. Control. Optim. *36* (1998), 929–947 (electronic). zbl MR
[34] *X. Y. Zhou, D. Li*: Continuous-time mean-variance portfolio selection: A stochastic LQ framework. Appl. Math. Optimization *42* (2000), 19–33. zbl MR

Authors' addresses: *Mokhtar Hafayed*, Laboratory of Applied Mathematics, P. O. Box 145, Biskra University 07000, Biskra, Algeria, e-mail: `hafa.mokh@yahoo.com`; *Petr Veverka*, Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Trojanova 13, Praha 120 00, Czech Republic, e-mail: `panveverka@seznam.cz`; *Syed Abbas*, School of Basic Sciences, Indian Institute of Technology Mandi, Mandi H.P. 175001 India, e-mail: `sabbas.iitk@gmail.com`.